

**UNIVERSIDADE DE LISBOA
FACULDADE DE LETRAS
DEPARTAMENTO DE LINGUÍSTICA GERAL E ROMÂNICA**



As Relações Filogenéticas entre os Crioulos do Golfo da Guiné

Abigail Tiny Cosme

**Dissertação de Mestrado em Linguística
Lisboa
2014**

**UNIVERSIDADE DE LISBOA
FACULDADE DE LETRAS
DEPARTAMENTO DE LINGUÍSTICA GERAL E ROMÂNICA**



**As Relações Filogenéticas entre os Crioulos do Golfo da Guiné
Abigail Tiny Cosme**

**Orientação: Professor Doutor Tjerk Hagemeijer
Co-orientação: Doutor Hugo Cardoso**

**Dissertação de Mestrado em Linguística
Lisboa
2014**

AGRADECIMENTOS

“Melhor é o fim das coisas do que o princípio delas”

Eclesiastes 7:8

“Melhor é a sabedoria do que a força”

Eclesiastes 9:16

Durante este longo percurso, muitos foram os que, direta e indiretamente, contribuíram para que eu pudesse dar mais este passo na incessante procura pelo conhecimento. Contudo, tudo teve um início. Agradeço ao meu país, São Tomé e Príncipe, por ter sido o meu berço e pela multiculturalidade de que sou fruto.

Agradeço ao projeto *“As origens e o desenvolvimento de sociedades crioulas no Golfo da Guiné: um estudo interdisciplinar”* (PTDC/CLE-LIN/111494/2009) por me ter aproximado dos crioulos de São Tomé e Príncipe e pelo seu contributo para a conservação e revitalização destas línguas.

É com grande reconhecimento que deixo um agradecimento especial aos meus orientadores. Ao Doutor Hugo Cardoso por ter aceitado prontamente o desafio de co-orientar esta tese. A partilha do seu conhecimento, a presença constante e a sua dedicação foram fundamentais. Ao Prof. Doutor Tjerk Hagemeijer por ter acreditado que seria possível, mesmo antes de mim. O tempo dispendido, a paciência demonstrada, a constante preocupação, a compreensão e a sua amizade tornaram a minha tarefa mais leve. Palavras não serão suficientes para agradecer.

À minha grande família Tiny espalhada por este mundo fora saibam que cada trilha que percorro e cada vitória que alcanço tem um pouco de todos vós.

À minha Crew... somos uma só em várias réplicas. Pelos dias bons mas, principalmente, pelos dias maus, obrigada meninas.

Aos meus irmãos por existirem. A vida é tão mais fácil convosco ao meu lado.

Aos meus pais por todos os dias me ensinarem o caminho da sabedoria, através do exemplo que são as suas vidas.

Ao Abdel e à Djamila pelas ausências perdoadas e acima de tudo por terem sido o meu porto seguro quando e sempre que precisei de refúgio.

Ao meu Deus por ser a razão de tudo.

Conteúdo

AGRADECIMENTOS	iii
RESUMO.....	v
ABSTRACT.....	vii
Lista de Abreviaturas	ix
Abreviaturas	ix
1. Introdução.....	1
1.1. Objetivos.....	1
1.1.1. Metodologia	1
1.2. Os Crioulos do Golfo da Guiné	4
2. Léxico	8
2.1. Introdução	8
2.2. Léxico básico	8
2.2.1. Introdução.....	8
2.2.2. Dados.....	9
2.2.3. Análise.....	13
2.3. Léxico Funcional	14
2.3.1. Introdução.....	14
2.3.2. Dados.....	14
2.3.3. Análise.....	18
2.4. Conclusão.....	19
3. Sintaxe	21
3.1. Introdução	21
3.1.1.1. Estudos filogenéticos anteriores	21
3.2. Dados Sintáticos.....	27
3.2.1. Análise.....	35
3.3. Conclusão.....	36
4. Conclusões finais	38
Referências Bibliográficas	41
ANEXOS	45
ANEXO I - Lista de 200 palavras de Swadesh: Origem do vocabulário.....	45
ANEXO II - Lista de 200 palavras Swadesh: Cognatos – Versão para <i>SplitsTree4</i>	50
ANEXO III - Lista de léxico funcional: Origem do vocabulário	55
ANEXO IV - Lista do léxico funcional: Classificação de cognatos.....	59
ANEXO V – Lista de traços sintáticos: Classificados.....	63
ANEXO VI - Lista de traços sintáticos – Versão para <i>SplitsTree4</i>	72

RESUMO

Esta tese tem como objetivo medir e avaliar distâncias entre os quatro crioulos do Golfo da Guiné e contribuir para o debate sobre a origem e o desenvolvimento destas línguas utilizando um novo método de análise comparativa. Assumindo a visão da generalidade dos estudiosos (e.g. Günther 1973; Ferraz 1979; Schang 2000; Hagemeijer 2011) de que os quatro crioulos do Golfo da Guiné –designadamente o Santome (falado na ilha de São Tomé), o Lung’le (falado na ilha do Príncipe), o Angolar (falado pela comunidade angolar da ilha de São Tomé) e o Fa d’Ambô (falado na ilha de Ano Bom) – constituem uma unidade genética que deriva de um antepassado comum, apresentar-se-á um estudo comparado do léxico e da sintaxe destas quatro línguas.

Em relação ao léxico, compilou-se uma lista de léxico básico de Swadesh (1955) de 200 palavras e uma lista de léxico funcional para cada um dos quatro crioulos; para a sintaxe, foi feito um levantamento de traços sintáticos constantes no *Atlas of Pidgin and Creole Language Structures* (APiCS) (Michaelis *et al.* 2013), complementado com uma recolha de novos traços relevantes para os crioulos do Golfo da Guiné.

Os dados recolhidos foram preparados e classificados para serem introduzidos no *SplitsTree4* (Huson & Bryant 2006), um programa de modelos estatísticos e de probabilidades que permitirá visualizar a distribuição dos dados através de redes filogenéticas. Com base nas redes geradas através dos dados lexicais e sintáticos recolhidos, calculou-se a distância a que as línguas se encontram entre si e a respetiva distância ao centro de traços partilhados. A análise destes resultados permitiu-nos verificar o grau de convergência e de divergência entre estas línguas.

Concluiu-se que o Santome, ao ocupar sistematicamente a posição de maior proximidade com o centro de dados partilhados, assume um papel de centralidade em relação aos restantes crioulos. Nessa perspetiva, os dados irão ao encontro da hipótese de que o Santome é a continuação do proto-crioulo no tempo e no espaço e que as restantes línguas terão ramificado a partir dele. Verifica-se também que há uma constante relação de proximidade a nível lexical entre o Santome e o Lung’le e que o Angolar se posiciona a uma considerável distância em relação aos demais crioulos do Golfo da Guiné. A proximidade entre o Santome e o Lung’le pode ser entendida pela relação de

proximidade e por contacto histórico continuado. O Angolar, embora falado na mesma ilha que o Santome, ocorre em posição periférica, o que corrobora a ideia de que esta terá sofrido um processo de relexificação relacionado com as línguas Bantu (Lorenzino 1998). A nível sintático, o facto de o Fa d'Ambô se encontrar numa posição mais periférica permite-nos estabelecer um paralelo com a hipótese defendida de que esta língua ficou isolada das restantes, tendo evoluído de forma independente.

Palavras-Chave: Crioulos do Golfo da Guiné; redes filogenéticas; léxico básico; léxico funcional; sintaxe.

ABSTRACT

This dissertation aims to measure and evaluate the distance between the four Gulf of Guinea creoles and contribute towards the debate of the origin and development of these languages, through the application of a new method of comparative analysis. Assuming the view of most scholars (e.g. Günther 1973; Ferraz 1979; Schang 2000; Hagemeyer 2011) that the four Gulf of Guinea creoles – namely Santome (spoken on the Island of São Tomé), Lung’le (spoken on the Island of Príncipe), Angolar (spoken by the angolar community of the Island of São Tomé), and Fa d’Ambô (spoken on the Island of Ano Bom) – constitute a genetic unit which derives from a common ancestor, we did a comparative study of the lexicon and syntax of these four languages.

With regard to the lexicon, we compiled a 200-word list of the basic lexicon by Swadesh (1955) and a list of functional lexicon for each of the four creoles; for syntax, we made a survey of syntactic features included in the *Atlas of Pidgin and Creole Language Structures* (APiCS) (Michaelis *et al.* 2013), complemented with a collection of new features which are relevant for the Gulf of Guinea creoles.

The data collected were systematized and classified in order to be fed to *SplitsTree4* (Huson & Bryant 2006), a program of statistic modeling and probabilities which allows for the visualization of data distribution in the form of phylogenetic networks. Based on the networks generated from the lexical and syntactic data collected, we calculated the distance between each language and their divergence from the pool of shared features. The analysis of these results allowed us to ascertain the degree of convergence and divergence between these languages.

We conclude that Santome, which systematically occupies the closest position to the pool of shared features, plays a central role in the relationship between these creoles. In this sense, the data align with the hypothesis that Santome is the continuation of the proto-creole in space and time and that the other languages constitute ramifications. We also observe a constant proximity, in lexical terms, between Santome and Lung’le and that Angolar shows considerable distance from the other Gulf of Guinea creoles. The closeness between Santome and Lung’le is attributed to proximity and sustained historical contact. Angolar, though spoken on the same island as Santome, occupies a

peripheral position, which corroborates the idea that it has undergone a process of relexification under the influence of Bantu languages (Lorenzino 1998). Syntactically, the fact that Fa d'Ambô is found in a peripheral position supports the hypothesis that this language became isolated from the others, having developed independently.

Keywords: Gulf of Guinea creoles; phylogenetic networks; basic lexicon; functional lexicon; syntax.

Lista de Abreviaturas

Abreviaturas	Significados
ALUSTP	Alfabeto Unificado para a Escrita das Línguas Nativas de S. Tomé e Príncipe
ANG	Angolar
APiCS	Atlas of Pidgin and Creole Language Structures
ART	Artigo
BAT	Crioulo da Batavia
CAS	Crioulo de Casamansa
CGG	Crioulos do Golfo da Guiné
CHACa	Cavite Chabacano
CHATe	Ternate Chabacano
CHAZa	Zamboanga Chabacano
CVBr	Crioulo de Cabo Verde da Brava
CVSa	Crioulo de Cabo Verde de Santiago
CVSV	Crioulo de Cabo Verde de São Vicente
DEM	Demonstrativo
DIU	Indo-Português de Diu
FA	Fa d'Ambô
FDA	Fa d'Ambô
GBK	Crioulo da Guiné-Bissau
GG	Golfo da Guiné
KOR	Korlai
LI	Lung'Ie
PaK	Papiá Kristang
PAL	Palenquero
PAP	Papiamento

PCGG	Proto-Crioulo do Golfo da Guiné
PL	Plural
POSS	Possesivo
PREP	Preposição
PRI	Principense/Lung'Ie
PROG	Progressivo
REL	Relativo
SLP	Português de Sri Lanka
SN	Sintagma Nominal
ST	Santome
STO	Santome
TMA	Tempo Modo Aspetto
WALS	World Atlas of Language Structures
1PL	Primeira Pessoa do Plural
3PL	Terceira Pessoa do Plural

1. Introdução

1.1. Objetivos

O objetivo desta dissertação é o de efetuar um estudo comparativo de forma a avaliar e analisar o grau de distância a que as quatro línguas crioulas do Golfo da Guiné (daqui em diante CGG) se encontram entre si. Estas línguas são o Santome (ST), falado na ilha de São Tomé, o Lung'le (LI), falado na ilha do Príncipe, o Angolar (ANG), falado pela comunidade angolar da ilha de São Tomé, e o Fa d'Ambô (FA), falado na ilha de Ano Bom.

A posição que tem sido defendida pela maioria dos autores é a de que há uma relação de parentesco entre os CGG (e.g. Günther 1973; Ferraz 1979; Schang 2000; Hagemeijer 2011). Estes admitem que do contato linguístico resultante do povoamento de S. Tomé se desenvolveu uma língua crioula de base lexical portuguesa, o proto-crioulo do Golfo da Guiné (PCGG), i.e., uma língua que se formou na ilha de São Tomé e que posteriormente se terá ramificado em quatro. Esta é também a posição que assumiremos neste estudo.

Com o objetivo de avaliar o grau de distanciamento entre os CGG, faremos uma comparação entre os quatro CGG, de forma a identificar pontos comuns e pontos de divergência nos domínios do léxico e da sintaxe.

1.1.1. Metodologia

De modo a identificar traços lexicais e sintáticos partilhados entre os CGG, procedemos a uma recolha de dados com base nos estudos existentes e em elicitación.

Para o léxico, numa primeira fase e com o objetivo de encontrar cognatos, iremos recorrer à lista de 200 palavras de Swadesh (1955) que criou uma lista do vocabulário básico de uma língua, i.e., o vocabulário mais comum e por isso mais resistente à mudança. Através da contagem de cognatos, torna-se possível avaliar a distância a que as

línguas se encontram. Neste estudo aplicaremos o método léxico-estatístico de forma a calcular percentagens de cognatos nos CGG. Para a classificação dos cognatos, Swadesh fez o levantamento de uma quantidade de itens lexicais com características específicas, a serem testadas num conjunto de línguas. Os itens adequados para constarem na lista terão de ser universais, não culturalmente específicos, cujo conceito seja facilmente identificável e que correspondam a termos simples na maior parte das línguas. Como resultado, Swadesh (1955) publicou uma lista de 200 itens correspondentes ao léxico básico das línguas, i.e, correspondente aos itens lexicais que apresentam maior taxa de retenção. Para este estudo, esta lista será aplicada aos quatro CGG.

Numa segunda fase, ainda relativa ao léxico, consideramos importante contrastar os resultados que surgirão da análise do léxico básico com os dados do léxico funcional das línguas em questão. Nesse sentido, iremos proceder a um levantamento, o mais exaustivo possível, dos itens funcionais destas línguas, itens estes pertencentes às classes fechadas, nomeadamente artigos, preposições, locuções prepositivas, pronomes, numerais, conjunções, complementadores, marcas de TMA, quantificadores, marcadores de negação e focalizadores. Nesta recolha, iremos igualmente considerar categorias relativas a características mais específicas destas línguas, tais como as partículas finais, os verbos mais funcionais nas estruturas de verbos seriais e as marcas de apresentação.

Para que este estudo comparativo se torne mais abrangente e permita uma leitura de resultados não apenas a nível do léxico mas também a nível da estrutura destas línguas, propomo-nos trabalhar questões relacionadas com a sintaxe, de modo a verificar qual o impacto que esta área da gramática terá no eventual relacionamento entre estes crioulos. Para o levantamento dos traços sintáticos, recorreremos ao *Atlas of Pidgin and Creole Language Structures* (APiCS) (Michaelis *et al.* 2013), que reúne um conjunto de dados sincrónicos comparativos relativos a estruturas gramaticais e lexicais de 76 pidgins, crioulos e outras línguas de contacto em todo o mundo. A base de dados do APiCS consiste em 130 traços estruturais transversais às diferentes áreas da gramática, tais como a fonologia, a morfologia, a sintaxe e o léxico. O facto de os dados se apresentarem em forma de mapa e de estes se encontrarem disponíveis online permite a visualização precisa da distribuição destas línguas no mundo e permite também um fácil, prático e imediato acesso aos dados. O APiCS foi inspirado no *World Atlas of Language*

Structures (WALS), com o qual partilha 48 traços adaptados. Os restantes traços são exclusivos do APiCS.

Tendo em conta que, para este trabalho, nos iremos focar nas questões sintáticas destes crioulos, de forma a reunir o maior conjunto de traços possível, procederemos, por um lado, à extração de todos os traços sintáticos do APiCS, e, por outro lado, recolheremos e analisaremos um conjunto de novos traços sintáticos. Para além do APiCS, uma boa parte dos dados recolhidos para esta dissertação, tanto a nível lexical como sintático, foram retirados dos corpora dos quatro crioulos do Golfo da Guiné disponíveis em linha em <http://alfclul.clul.ul.pt/CQPweb/>, os quais abarcam um grande conjunto de dados orais e escritos relativos aos quatro CGG. Estes corpora foram recolhidos no âmbito do projeto “*As origens e o desenvolvimento das sociedades crioulas do Golfo da Guiné: um estudo interdisciplinar* (PTDC/CLE-LIN/111494/2009)”, financiado pela Fundação para a Ciência e Tecnologia (FCT). Após a recolha e tratamento dos dados mencionados anteriormente, iremos introduzi-los num programa de redes filogenéticas.

As redes filogenéticas são um instrumento criado na área da biologia (e, durante muito tempo, utilizado exclusivamente neste domínio) para identificar e classificar espécies. Contudo, esta ferramenta tem vindo a evidenciar-se na área dos estudos de evolução linguística, quer em estudos lexicais como estruturais, pois, através desta, torna-se possível estabelecer mais objetivamente famílias de línguas e analisar traços comuns ou divergentes entre línguas. Neste estudo, iremos utilizar a ferramenta *SplitsTree* v.4.11.3 (Huson & Bryant 2006). Trata-se um programa de modelos estatísticos e de probabilidades que permitirá visualizar as distâncias entre línguas numa estrutura em forma de rede.

Nos que diz respeito aos CGG, ao longo do tempo têm surgido estudos descritivos destas línguas (Ferraz 1979; Günther 1973; Lorenzino 1998; Maurer 1995, 2009; Schang 2000) e sobre a forma como se relacionam (e.g. Hagemeijer 2011; Maurer 1999). Porém, neste momento, verifica-se a inexistência de estudos comparados alargados. Com este estudo, pretendemos contribuir para a consolidação, com base em dados sistematizados, de propostas que têm sido feitas e defendidas sobre estes crioulos, mostrando tendências e levantando questões sobre as relações diacrónicas entre estas línguas.

Em relação à escrita destes crioulos, ao longo deste estudo, iremos adotar o Alfabeto Unificado para a Escrita das Línguas Nativas de S. Tomé e Príncipe (ALUSTP), que é, desde 2013, o alfabeto oficializado para o ST, o ANG e o LI, promulgado pelo decreto de lei nº19/2013 a 14 de agosto de 2013, no Diário da República de São Tomé e Príncipe. O ALUSTP pretende uniformizar e sistematizar a escrita das línguas nativas de São Tomé e do Príncipe utilizando, ao máximo, um princípio de correspondência biunívoca entre fonema e grafema. Como o FA é uma língua da Guiné Equatorial, não foi contemplada por esta escrita. Contudo, seguimos o mais possível o mesmo princípio para a transcrição dos dados desta língua.

1.2. Os Crioulos do Golfo da Guiné

Segundo os dados históricos, as ilhas de São Tomé e Príncipe foram descobertas nos anos de 1471/2 (Caldeira 2006) e a de Ano Bom entre 1483 e 1501 (Caldeira 2010). Das três ilhas, a primeira a ser povoada de forma permanente foi a de São Tomé, em 1493. Houve, no entanto, uma primeira tentativa de povoamento, menos bem sucedida, em 1485 (Ferraz 1976), nesta que é a ilha maior e centro das atividades económicas no Golfo da Guiné. A ilha do Príncipe terá sido povoada por volta do ano 1500 aquando da sua doação, por decreto real. A doação da ilha de Ano Bom ocorreu em 1503; porém, o seu povoamento definitivo parece ter ocorrido entre 1543 e 1565 (Caldeira 2006). Os Angolares têm sido descritos como descendentes de escravos quilombolas originários do primeiro povoamento de São Tomé (Caldeira 2004; Ferraz 1974; Lorenzino 1998; Seibert 2007). As fugas de escravos estão largamente documentadas desde as fases iniciais da habitação e tornaram-se uma séria ameaça para a povoação e as plantações a partir de 1530.

A ocupação de São Tomé pode ser dividida em duas fases: a fase de habitação e a fase de plantação. A primeira fase corresponde ao período entre 1485, data do primeiro povoamento, e cerca de 1515, altura que marca o início do ciclo do açúcar. A segunda fase vai até ao final do século XVI (e.g. Sousa 1990). O colapso da economia de plantação deu-se por volta do ano de 1600, como consequência da melhor qualidade do açúcar do Brasil e dos sucessivos ataques de escravos quilombolas às plantações.

A história da primeira fase está diretamente relacionada com o comércio de escravos no delta do Níger e particularmente com o antigo Reino de Benim, situado na atual Nigéria (Hagemeijer 2011). Há um consenso generalizado de que o número de escravos de origem Bantu aumentou significativamente depois de 1520, período que correspondeu ao rápido declínio do comércio do delta do Níger, que foi definitivamente cortado em meados do século XVI.

O contacto entre os portugueses e os escravos resgatados do delta do Niger, sobretudo do antigo Reino de Benim, na fase de habitação da ilha criou as condições ideais para a formação de um pidgin. Neste período, havia um contacto intenso e direto entre os portugueses e os escravos africanos do delta do Níger, que seriam sobretudo falantes do Edo, uma língua da família Edóide. Este pidgin terá sido rapidamente nativizado entre as gerações seguintes de escravos, tendo dado origem ao que designamos por proto-crioulo do Golfo da Guiné (PCGG) (Hagemeijer 2009, 2011). Esta língua terá ramificado no tempo e no espaço, dando origem aos quatro CGG.

Há um conjunto de estudos que dão conta da relação de proximidade existente entre os CGG. Matos (1842) foi o primeiro autor a fazer referência à relação entre três dos quatro CGG. O autor afirma que o LI e o ST são bastante semelhantes, diferindo apenas na quantidade de vocabulário africano, tendo o LI mais vocabulário africano. Quanto ao FA, o autor defende que é semelhante ao ST, diferenciando-se deste pela pronúncia gutural que apresenta, o que certamente remete para a presença da fricativa velar /x/ que não existe nos outros CGG. Valkhoff (1966) partilha desta mesma visão, afirmando que o ST, o LI e o FA são dialetos de uma mesma língua. No entanto, esta afirmação do autor pode levar a uma interpretação de que estas línguas seriam mutuamente inteligíveis, o que não se verifica.

Durante muito tempo, o ANG foi um crioulo relativamente desconhecido e associado às línguas Bantu. Negreiros (1895) entendia o ANG como uma mistura entre o ST e o Kimbundu e Valkhoff (1966) chegou a afirmar que esta seria uma língua Bantu. Ferraz (1974) foi o primeiro a identificar o ANG como uma língua crioula e estudos posteriores (Maurer 1992, 1995; Lorenzino 1998) confirmaram esta hipótese, concluindo que a componente Bantu está sobretudo relacionada com a presença maciça de léxico desta família.

Nos seus estudos, Ferraz aborda duas visões diferentes em relação a estes crioulos. Por um lado, o autor argumenta a favor de uma relação genética entre os quatro CGG, com base numa proto-língua, i.e., defende a existência de um proto-crioulo que teria sido o ST original e que mais tarde ramificaria em quatro crioulos, por via de separação geográfica e também de eventuais diferenças a nível do substrato (Ferraz 1976). Esta visão foi, no entanto, por si abandonada mais tarde, quando afirma:

“To take the GG [Gulf of Guinea] case, it would not be plausible to assume that the contact language which developed in the town of São Tomé and the surrounding areas was the same as that which gave rise to Ang[olar], Pr[incipense], and Pag[alu]. There are enough differences between each of these languages to rule out such a possibility. It would be closer to the truth to say that the four contact languages show many resemblances because, to a large extent, they grew up together, with slaves and settlers introduced through the central administration in São Tomé ... Hence different languages developed in the archipelago rather than dialects of one contact language.” (Ferraz 1987b: 348)

O autor considera, portanto pouco plausível que o contato linguístico que terá ocorrido na cidade de São Tomé tenha sido o mesmo a dar origem aos restantes crioulos (por Pagalu entenda-se Fa d’Ambô), afastando-se da hipótese de uma origem genética comum. Contudo, convém não esquecer que o trabalho de Ferraz foi maioritariamente direcionado para o ST, com maior ênfase em questões de léxico e fonologia. Os numerosos trabalhos sobre os CGG por diferentes autores, a partir de 1990 (incluindo gramáticas, teses e artigos,) vieram a reforçar a hipótese de uma origem comum (ver Hagemeijer 2011).

Do ponto de vista linguístico, as línguas do delta do Níger da primeira fase tiveram visível impacto tipológico enquanto estrato africano primário na formação do PCGG. As estratégias de serialização verbal, de reflexivização com a palavra para corpo, a existência de implosivas e de lábio-velares são alguns dos exemplos que demonstram esse impacto. A fase de plantação, com a chegada maciça de escravos Bantu em detrimento de escravos do delta do Níger, contribuiu para a afirmação do papel do Bantu

ocidental, nomeadamente do Kikongo e do Kimbundu, nestes crioulos, com especial incidência a nível do léxico e da fonologia (Hagemeijer 2009). Em relação à influência das línguas africanas nos CGG, os diferentes estudos têm apontado para o facto de o LI e o ANG terem um menor grau de relação no que diz respeito ao léxico africano, devido à significativa influência lexical do Edo e do Kimbundu nas respetivas línguas. O ST e o FA, por sua vez, evidenciam uma quantidade semelhante de léxico quer do Edo quer do Kikongo. As claras diferenças que hoje se podem observar nos quatro CGG são resultado do relativo isolamento em diferentes momentos da história, do efeito de contato posterior e do papel da mudança interna.

Em suma, grande parte dos estudiosos sobre esta área defende, generalizadamente, que os quatro CGG partilham uma origem comum (Hagemeijer 1999, 2009; Lorenzino 1998; Maurer 1995, 1999; Schang 2000) e este será também o ponto de partida deste estudo.

2. Léxico

2.1. Introdução

Os estudos do léxico têm sido importantes no contexto das investigações linguísticas para a análise, comparação e classificação de línguas. Neste capítulo, analisaremos questões relacionadas com o léxico das quatro línguas crioulas em estudo, a saber, Santome, Lung'Ie, Angolar e Fa d'Ambô. Na secção 2.2., faremos uma análise do léxico básico com base na lista de 200 palavras de Morris Swadesh (1955) e, de seguida, na secção 2.3., analisaremos igualmente o léxico funcional com base numa lista exaustiva de dados recolhidos especificamente para este estudo. Deste modo, pretende-se avaliar a distância a cada língua se encontra entre si com base nos diferentes dados. Na secção 2.4., estabeleceremos uma comparação entre o resultados das duas secções anteriores.

2.2. Léxico básico

2.2.1. Introdução

O léxico básico tem desempenhado um papel importante na comparação e classificação de línguas em famílias linguísticas, i.e. para determinar se um determinado conjunto de línguas deriva de uma proto-língua comum, ao aplicar-lhe métodos de reconstrução, em particular o Método Comparativo. Neste estudo, utilizamos a lista Swadesh extensa, composta por 200 itens lexicais básicos. Este léxico caracteriza-se pela sua universalidade, não denotando, portanto, especificidades culturais. Estes itens apresentam também uma maior resistência à mudança linguística. A comparação quantitativa de cognatos, conhecida como método léxico-estatístico (e.g. Swadesh 1955), permite-nos postular hipóteses sobre o grau de proximidade entre línguas através das percentagens de cognatos partilhados. O nosso objetivo aqui é indicar as relações de cognacia no léxico básico e gerar uma rede filogenética que mostre as distâncias entre os quatro CGG.

Em relação aos CGG, Ferraz (1979) foi o primeiro a comparar as percentagens de léxico partilhado entre o ST e os demais crioulos, mostrando que o ST partilha 88% do léxico com o LI, 82% com o FA e 67% com o ANG. Quanto à origem deste léxico, é maioritariamente de origem portuguesa, embora tenha sofrido alterações fonológicas e

semânticas significativas. O léxico africano constitui uma pequena percentagem do léxico total. As mais importantes diferenças entre o léxico destes crioulos registam-se a nível do léxico africano, onde se encontram itens lexicais provenientes dos dois mais importantes grupos linguísticos de substrato: o Edóide e o Bantu.

2.2.2. Dados

A lista de 200 palavras de Swadesh foi aplicada aos 4 CGG (Anexo I), com recurso a dicionários (Araújo & Hagemer 2013), gramáticas (Maurer 1995, 2009; Zamora 2010) e dados elicitados com falantes nativos. O processo de classificação e numeração dos cognatos foi feito manualmente. Procedeu-se à comparação de cada uma das entradas nos quatro CGG de forma a detetar elementos comuns que permitissem identificar possível cognacia entre o vocabulário básico. Após este processo, fez-se a devida conversão para o programa *SplitsTree4* de forma a verificar, através da rede filogenética criada, a que distância estas línguas se encontram entre si e qual a relação entre elas.

A Tabela 1 a seguir mostra um exemplo de palavras cognatas.

Português	Santome	Lung'le	Angolar	Fa d'Ambô
preto	<i>pletu</i>	<i>peetu</i>	<i>peetu</i>	<i>peetu</i>
erguer	<i>ligi</i>	<i>rêgê</i>	<i>rigi/digi</i>	<i>ligi</i>

Tabela 1. Palavras cognatas de origem portuguesa

De acordo com os exemplos da Tabela 1, é possível verificar que as palavras identificadas são palavras cognatas pois revelam-se semelhantes nas quatro línguas. As diferenças existentes são justificáveis através de processos fonético-fonológicos regulares que terão ocorrido ao longo da formação destas línguas, tais como a manutenção da líquida /r/ no LI e o alongamento compensatório de vogais (Ferraz 1987a). Por outro lado, foi também possível identificar exemplos em que, em uma ou mais línguas, a origem etimológica da palavra não coincidia. Nesses casos, poder-se-á conjecturar que a

entrada de um determinado item lexical se pode dar de diferentes formas, por exemplo a partir de um vocábulo diferente de uma mesma língua ou a partir de vocábulos de línguas diferentes. Em ambos os casos, não existe cognacia. Na Tabela 3, encontram-se dados da lista Swadesh (Anexo 1) de onde se podem retirar exemplos de divergência de origem etimológica. Repare-se, por exemplo, que a palavra *tata* para ‘pai’ em ANG diverge das restantes línguas, nas quais os vocábulos – *pe* (ST), *pwe* (LI) e *pe* (FA) – são cognatos,.

Português	Santome	Lung'ŋe	Angolar	Fa d'Ambô
pai	<i>pe</i>	<i>pwe</i>	<i>tata</i>	<i>pe</i>
peixe	<i>pixi</i>	<i>pêxi</i>	<i>kikiê</i>	<i>pixi</i>
pescoço	<i>klonko</i>	<i>ôô</i>	<i>godxi</i>	<i>khôkhôndjô</i>

Tabela 2. Léxico de base etimológica não comum

Após análise e classificação dos dados, estes foram introduzidos num programa de leitura de modelos estatísticos e de probabilidades de forma a identificar um padrão, o que permitirá o levantamento de hipóteses. O programa gera redes de proximidade (Figura 1) e redes com raiz (Figura 2) que indicam a distância entre as línguas.

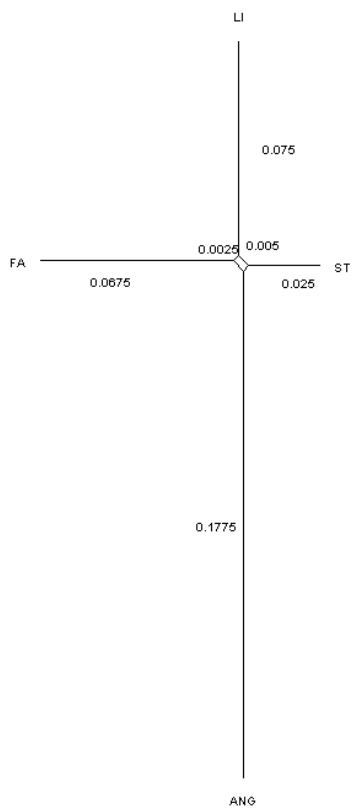


Figura. 1 Lista de Swadesh aplicada à Rede de Proximidade (Neighbournet)

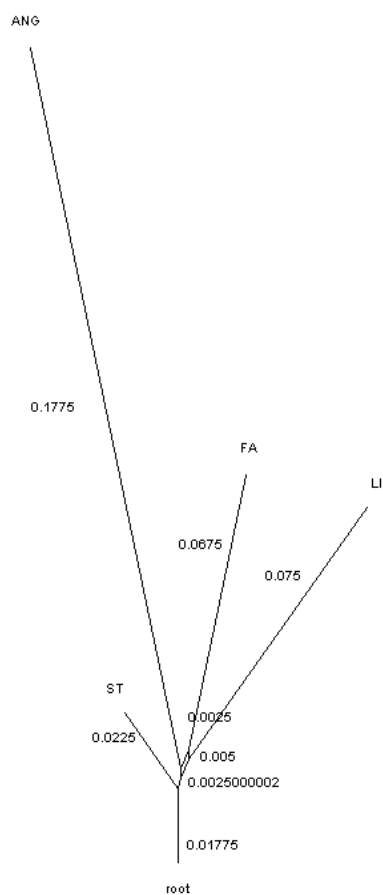


Figura. 2 Lista de Swadesh aplicada à Rede com Raiz (*Rooted Network*)

As distâncias numéricas entre as línguas nas redes acima representadas encontram-se hierarquizadas na Tabela 3, ordenadas do par mais próximo para o mais distante.

Pares	Distância
ST – FA	0.1
ST – LI	0.105
LI - FA	0.145
ST – ANG	0.205
ANG – FA	0.25
LI – ANG	0.26

Tabela 3. Distância numérica entre os CGG – Léxico básico

Para que melhor se entenda a distribuição e o grau de proximidade dos itens lexicais, fez-se a contagem para a lista de Swadesh de forma a calcular percentagens relativamente à origem das palavras nos 4 CGG. A tabela que se segue apresenta o resultado desse estudo.

	ST	LI	ANG	FA
Português	92,00%	91,00%	75,50%	91,50%
Africano	5,00%	5,50%	10,00%	5%
Outro/desconhecido	3,00%	3,50%	14,50%	3,50%

Tabela 4. Origem do léxico na lista de Swadesh em percentagens

2.2.3. Análise

Ao analisar a rede da Figura 1, verifica-se que o ST é o que se encontra mais próximo do centro de partilha. O ANG, por sua vez, é o que se encontra a uma maior distância do centro e das restantes línguas.

O centro da figura representa o centro de traços partilhados. A dimensão do centro é proporcional à proximidade entre as línguas, i.e., quanto menor for o centro, maior é o número de traços que partilham e menor é a distância entre as línguas. A Figura 1 revela que o centro de traços partilhados é reduzido e bastante uniforme. Poder-se-á, deste modo, inferir que os 4 CGG partilham um grande número de léxico básico.

Os dados mostram ainda que o ST é o que se encontra claramente mais próximo do centro (0,025), seguido do FA (0,0675) e do LI (0,075), encontrando-se estas três línguas mais próximas, em termos de distância, entre si. No entanto, apesar de o ANG se encontrar a uma maior distância, este diverge no mesmo sentido que o ST, encontrando-se os dois no mesmo vértice da rede.

Na Figura 2 está representada a relação de proximidade dos quatro crioulos à sua raiz, sendo que o ponto de partida desta rede (a raiz) corresponde ao conjunto de dados partilhados. A partir dela, a rede ramifica-se de acordo com as características que

divergem nas diferentes línguas. Nesta rede, tal como na anterior, pode-se observar a maior proximidade do ST, LI e FA à raiz. Através desta, torna-se ainda mais clara a tendência para a bipolarização do ST e ANG, por um lado, e do LI e FA, por outro. O resultado da análise do centro de traços/itens lexicais partilhados e da raiz, nas Figuras 1 e 2, vêm corroborar hipóteses já defendidas pelos diferentes autores sobre a origem genética comum destes crioulos (Ferraz 1979; Schang 2000; Hagemeyer 2011, entre outros).

2.3. Léxico Funcional

2.3.1. Introdução

A par do léxico básico, o léxico funcional poderá contribuir para apontar algumas respostas em relação ao comportamento dos CGG. Nesta secção, focar-se-ão aspectos relacionados com o léxico funcional, tendo por base uma lista exhaustiva dos itens lexicais funcionais das quatro línguas crioulas, aplicada a uma rede de proximidade.

2.3.2. Dados

Nas línguas, existem classes abertas e classes fechadas cujas principais diferenças se encontram na permeabilidade e resistência à mudança. As classes abertas caracterizam-se pelo facto de serem constituídas por um número potencialmente ilimitado de vocábulos, sendo possível o acréscimo de novas palavras no decurso da evolução da língua. São exemplos de classes abertas em português os nomes, verbos, adjetivos e advérbios. Por outro lado, as classes fechadas são compostas por um número limitado de palavras, correspondentes a itens cuja função é gramatical. Nestas classes, dificilmente são adicionados novos vocábulos e delas fazem parte, por exemplo, pronomes, preposições, conjunções, determinantes e quantificadores.

Sendo o léxico funcional considerado menos permeável a mudanças, afigura-se interessante estudar o seu comportamento nos 4 CGG, de forma a verificar se este se mantém, efetivamente, próximo nas diferentes línguas e se este apresenta a mesma tendência que o léxico básico, onde se pode encontrar um conjunto de itens lexicais e

funcionais considerados mais frequentes nas línguas. Nesse sentido, foi feito o levantamento exaustivo das unidades funcionais das línguas em análise (Anexo III). Para este estudo, fez-se o levantamento das seguintes categorias: artigos, preposições, locuções prepositivas, pronomes, numerais, conjunções, complementadores, marcas de TMA, partículas, quantificadores, marcadores de negação, focalizadores, verbos mais funcionais nas estruturas de verbos seriais e marcas de apresentação. Das diferentes categorias, procurou fazer-se um levantamento exaustivo, apesar de condicionado aos dados existentes e disponíveis destas línguas. A Tabela 5 é composta por itens retirados da lista de Léxico Funcional. Foi selecionada para exemplo uma classe de itens funcionais, os pronomes pessoais, aqui apresentada com a sua respetiva classificação numérica em termos de cognacia.

Pronomes Pessoais		Santome		Lung'Ie		Angolar		Fa d'Ambô
1ª sg	<i>l</i>	<i>n</i>	<i>l</i>	<i>n</i>	<i>l</i>	<i>n</i>	<i>l</i>	<i>am</i>
2ª sg	<i>l</i>	<i>bô</i>	<i>2</i>	<i>txi</i>	<i>l</i>	<i>bô</i>	<i>l</i>	<i>bo</i>
3ª sg	<i>l</i>	<i>ê</i>	<i>l</i>	<i>ê</i>	<i>l</i>	<i>ê</i>	<i>l</i>	<i>ê</i>
Impessoal	<i>l</i>	<i>a</i>	<i>l</i>	<i>a</i>	<i>l</i>	<i>a</i>	<i>l</i>	<i>a</i>
1ª pl	<i>l</i>	<i>non</i>	<i>l</i>	<i>non</i>	<i>l</i>	<i>no</i>	<i>l</i>	<i>non</i>
2ª pl	<i>l</i>	<i>inansê</i>	<i>2</i>	<i>owo</i>	<i>3</i>	<i>êthê</i>	<i>l</i>	<i>namsêji</i>
3ª pl	<i>l</i>	<i>inen</i>	<i>l</i>	<i>inen</i>	<i>l</i>	<i>ane</i>	<i>l</i>	<i>êneyn</i>

Tabela 5. Pronomes pessoais de sujeito em posição argumental.

Após levantamento e classificação dos itens em cognatos e não cognatos, procedeu-se de modo idêntico ao descrito para o léxico básico. Os dados foram introduzidos em ficheiros e posteriormente analisados e convertidos pelo programa *SplitsTree4*, em estruturas de redes, através de modelos estatísticos e de probabilidades (Figuras 3 e 4).

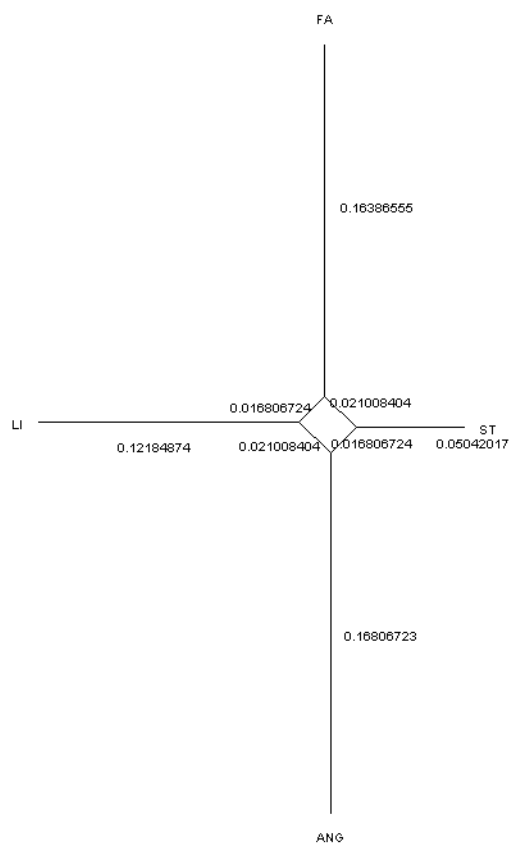


Figura 3. Lista de léxico funcional aplicada à Rede de Proximidade (*Neighbournet*)

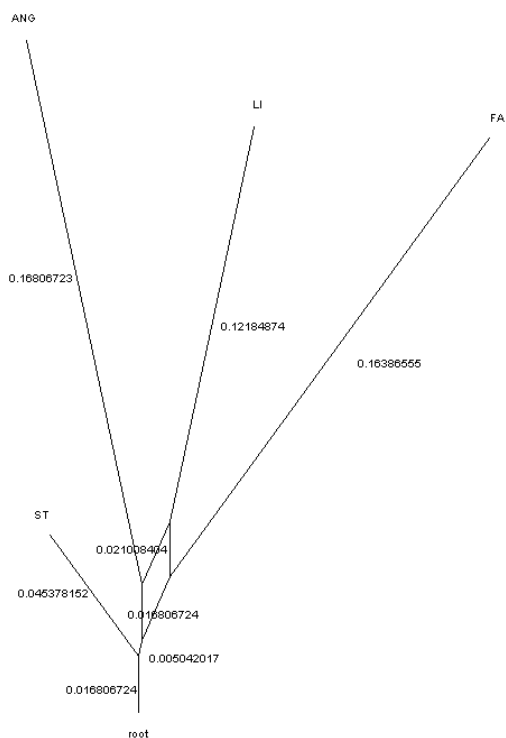


Figura 4. Lista de léxico funcional aplicada à Rede com Raiz (*Rooted Network*)

Tal como no caso do léxico básico, as distâncias numéricas entre as línguas nas redes acima representadas foram ordenadas do par mais próximo para o mais distante, como mostra a seguinte tabela.

Pares	Distância
ST – LI	0.21
ST – ANG	0.235
ST – FA	0.235
LI – FA	0.30
LI - ANG	0.31
ANG - FA	0.369

Tabela 6. Distância numérica entre os CGG - léxico funcional

Por fim, à semelhança da Tabela 4 acima, incluímos uma tabela que apresenta as percentagens do léxico funcional por origem.

	ST	LI	ANG	FA
Português	75,00%	80,00%	66,00%	76,00%
Africano	15,00%	13,00%	22,00%	11%
Outro/desconhecido	10,00%	7,00%	12%	13,00%

Tabela 7. Origem do léxico funcional na lista de Swadesh em percentagens

2.3.3. Análise

Ao analisar a Figura 3, é possível verificar que o centro da Figura, correspondente aos traços partilhados, é de maior dimensão quando comparado com o centro da Figura 1. Este dado revela que a proporção de itens funcionais partilhados entre os quatro CGG é menor, o que faz com que haja uma maior dispersão da informação e consequente dilatação do centro.

O ST é o que se apresenta mais próximo do centro de partilha, bem como mais próximo da raiz, conforme se pode verificar nas Figuras 3 e 4. Esta proximidade torna-se particularmente evidente na rede da Figura 4, em que, em comparação com as outras línguas, o ST está claramente menos distante da raiz, a uma distância de 0,045. O ANG, por sua vez, é o que se encontra mais distante do centro nas duas Figuras. No entanto, quando comparado com o FA, a diferença em termos de distância, tanto em relação ao centro de dados partilhados como à raiz, é bastante reduzida. Contudo, apesar do claro desfasamento entre o ST e o ANG em termos de distância às bases, os dois encontram-se no mesmo vértice na rede da Figura 3 e numa maior relação de proximidade na Figura 4. O desfasamento em relação ao ANG pode ser justificado pela elevada percentagem de léxico africano que este contém (cf. Tabela 7)

2.4. Conclusão

O centro de dados partilhados na Figura 1 é proporcionalmente mais pequeno em relação ao centro da Figura 3. Este dado revela que há uma maior partilha de itens entre todas as línguas, i.e., existe uma quantidade superior de dados comuns aos quatro crioulos, no léxico básico do que no léxico funcional.

Nas quatro representações de redes, o ST é o que se apresenta consistentemente mais próximo do centro e da raiz, o que revela que este é o que mais informação partilha com as restantes línguas e demonstra a sua centralidade em relação aos restantes crioulos.

O facto de o LI ter sofrido pouca influência das línguas Bantu (Hagemeijer 2011) pode permitir a leitura de um isolamento do LI numa extremidade da rede, em detrimento das restantes línguas. Por outro lado, no que diz respeito ao léxico funcional, é possível justificar esta troca de posição do LI com o FA na medida em que o FA, após a separação do proto-crioulo, deixou de ter contacto/influência com as restantes línguas, ao passo que o ST, o LI e o ANG permaneceram em contacto ao longo do tempo. Neste sentido, o FA na Figura 3 poderá estar em posição de isolamento em relação aos restantes três crioulos.

O ANG é o que se apresenta sistematicamente mais distante do centro de dados partilhados, bem como das restantes línguas. A influência e o impacto das línguas Bantu no ANG (Lorenzino 1998, Hagemeijer 2011), relacionados com a fase de plantação, justifica este maior distanciamento da língua. Este impacto é corroborado pelas Tabela 4 e 7, que demonstram que há claramente maior impacto das línguas africanas no léxico básico e funcional do ANG. No entanto, o ANG apresenta consistentemente uma grande relação de proximidade com o ST, sendo que nas Figuras 1 e 3 ocorrem no mesmo vértice e nas Figuras 2 e 4 em ramificações mais próximas. Tendo em conta que o ST e o ANG permaneceram na mesma ilha, poderá ter existido contacto continuado que eventualmente terá aproximado e ditado características específicas dos crioulos falados na ilha de S. Tomé. Nesse sentido, a proximidade entre estas duas línguas pode justificar o facto de, aquando da divergência das quatro línguas, o ST e o ANG divergirem na mesma direção.

Já foi dito anteriormente que as unidades funcionais são as mais estáveis e menos passivas de mudança. A análise às Figuras permite-nos aferir que há, de facto, uma maior

proximidade entre as línguas a nível do léxico funcional; porém, a nível do léxico básico, o distanciamento do ANG em relação ao centro de partilha nas Figuras 1 e 2 consolida a hipótese de que as línguas Bantu tiveram impacto a nível do léxico básico, tendo possivelmente alterado e substituído léxico crioulo original. Estes dados corroboram a hipótese da relexificação de Lorenzino (1998).

3. Sintaxe

3.1. Introdução

Para este estudo, consideramos relevante analisar questões relacionadas com a sintaxe destas línguas para que melhor se perceba de que forma o seu funcionamento determinará o grau de proximidade ou de distância entre elas. Para tal, iremos recorrer a um conjunto de traços sintáticos que abarcam traços mais gerais selecionados do APiCS e traços mais específicos por nós elicitados. Tal como já foi dito anteriormente, iremos extrair traços sintáticos que se encontram no APiCS pois este disponibiliza uma quantidade significativa de dados organizados e sistematizados.

Após a fase de recolha e tratamento dos dados, iremos introduzi-los no programa das redes filogenéticas, o que permitirá visualizar a distribuição e distanciamento entre os CGG através das diferentes redes que o programa permite gerar. Deste modo, utilizaremos estas redes para classificar uma família linguística já identificada e descrita na literatura (Hagemeijer 2009a, 2011; Maurer 1995, 1999; Schang 2000).

Este capítulo será dividido em duas partes. Numa primeira secção, iremos referenciar dois estudos anteriores que apresentam uma abordagem semelhante à que nos propomos, de forma a estabelecermos ligações e possíveis comparações no processo de análise dos dados e dos resultados obtidos. Apresentaremos estudos que fazem uma abordagem de cariz tipológico tais como Bakker *et al.* (2011), que fez uso das ferramentas filogenéticas com o intuito de provar que os crioulos formariam um grupo independente das restantes línguas do mundo. Apresentaremos ainda um outro estudo focado em aspetos estruturais, nomeadamente Sippola & Tiny (2014), que, através de uma abordagem sintática, procurou avaliar distâncias entre os crioulos ibéricos.

3.1.1.1. Estudos filogenéticos anteriores

Os estudos que utilizam redes filogenéticas para analisar relações entre a estrutura de línguas têm vindo a aumentar. McMahon & McMahon (2003) argumentam que a classificação das línguas deve ser quantificada de forma a permitir uma avaliação

objetiva dos métodos linguísticos alternativos. Para que isso aconteça, o primeiro passo passa pela adoção das ferramentas computacionais da biologia.

Um exemplo da utilização destas ferramentas está patente no estudo de Bakker *et al.* (2011). No sentido de tentar atribuir às línguas crioulas um estatuto diferente das restantes línguas do mundo e reuni-las num grupo tipológico distinto e único, Bakker *et al.* (2011) apresentam um estudo em que comparam traços morfossintáticos de línguas crioulas e não crioulas. Para interpretar estes traços, são usados modelos estatísticos e ferramentas computacionais de tipologia quantitativa, nomeadamente redes e árvores filogenéticas.

Um outro exemplo é o estudo de Sippola & Tiny (2014) que analisa a sintaxe nominal dos crioulos ibéricos das diferentes áreas, com diferentes línguas de substrato. A amostra para o estudo inclui tanto as variantes atlânticas como as asiáticas que se encontram na base de dados do APiCS. Para este estudo foram seleccionadas 19 línguas crioulas de base ibérica nas quais foram analisados 21 traços relacionados com a sintaxe nominal. A Tabela abaixo discrimina as línguas usadas para este estudo comparativo.

Língua	Abreviatura	Língua de Superstrato	Área
<i>Crioulo de Cabo Verde de Santiago</i>	<i>CVSa</i>	<i>Português</i>	<i>Alta Guiné</i>
<i>Crioulo de Cabo Verde da Brava</i>	<i>CVBr</i>	<i>Português</i>	<i>Alta Guiné</i>
<i>Crioulo de Cabo Verde de São Vicente</i>	<i>CVSV</i>	<i>Português</i>	<i>Alta Guiné</i>
<i>Crioulo da Guiné-Bissau- Kriol</i>	<i>GBK</i>	<i>Português</i>	<i>Alta Guiné</i>
<i>Crioulo de Casamansa</i>	<i>CAS</i>	<i>Português</i>	<i>Alta Guiné</i>
<i>Santome</i>	<i>STO</i>	<i>Português</i>	<i>Golfo da Guiné</i>
<i>Angolar</i>	<i>ANG</i>	<i>Português</i>	<i>Golfo da Guiné</i>
<i>Principense</i>	<i>PRI</i>	<i>Português</i>	<i>Golfo da Guiné</i>
<i>Fa d'Ambô</i>	<i>FDA</i>	<i>Português</i>	<i>Golfo da Guiné</i>
<i>Indo-Português Diu</i>	<i>DIU</i>	<i>Português</i>	<i>Ásia do Sul</i>
<i>Korlai</i>	<i>KOR</i>	<i>Português</i>	<i>Ásia do Sul</i>
<i>Português Sri Lanka</i>	<i>SLP</i>	<i>Português</i>	<i>Ásia do Sul</i>
<i>Papiá Kristang</i>	<i>PaK</i>	<i>Português</i>	<i>Sudeste Asiático</i>
<i>Crioulo da Batavia</i>	<i>BAT</i>	<i>Português</i>	<i>Sudeste Asiático</i>
<i>Ternate Chabacano</i>	<i>CHATe</i>	<i>Espanhol</i>	<i>Filipinas</i>
<i>Cavite Chabacano</i>	<i>CHACa</i>	<i>Espanhol</i>	<i>Filipinas</i>
<i>Zamboanga Chabacano</i>	<i>CHAZa</i>	<i>Espanhol</i>	<i>Filipinas</i>
<i>Papiamento</i>	<i>PAP</i>	<i>Espanhol</i>	<i>Caraíbas</i>
<i>Palenquero</i>	<i>PAL</i>	<i>Espanhol</i>	<i>Caraíbas</i>

Tabela 8. Crioulos ibéricos utilizados no estudo dos traços nominais (Sippola & Tiny 2014).

A Tabela que se segue apresenta a numeração dos traços de acordo com a disposição do APiCS.

<i>2. Ordem do possuidor e do possuído</i>
<i>3. Ordem do adjetivo e do nome</i>
<i>4. Ordem da adposição e do sintagma nominal</i>
<i>5. Ordem do demonstrative e do nome</i>
<i>6. Ordem do cardinal numeral e do nome</i>
<i>7. Ordem da oração relativa e do nome</i>
<i>9. Posição do artigo definido no sintagma nominal</i>
<i>10. Posição do artigo indefinido no sintagma nominal</i>
<i>22. Ocorrência de marcadores nominais de plural</i>
<i>23. Expressão de significado nominal plural</i>
<i>28. Artigos definidos</i>
<i>29. Artigos Indefinidos</i>
<i>30. Sintagmas nominais genéricos em função de sujeito</i>
<i>31. Co-ocorrência de demonstrativos e artigo definido</i>
<i>32. Demonstrativos pronominais e adnominais</i>
<i>33. Contrastes de distância nos demonstrativos</i>
<i>34. Numerais adnominais distributivos</i>
<i>37. Marcadores de possuidores pronominais</i>
<i>38. Marcador de sintagma nominal possuidor</i>
<i>39. Possuidores pronominais independentes</i>
<i>40. Concordância de género de adjetivos adnominais</i>

Tabela 9. Traços nominais dos crioulos ibéricos (Sippola & Tiny 2014)

O resultado da análise dos 21 traços correspondentes à sintaxe nominal das línguas crioulas de base lexical portuguesa e espanhola é dado na Figura 5, abaixo indicada.

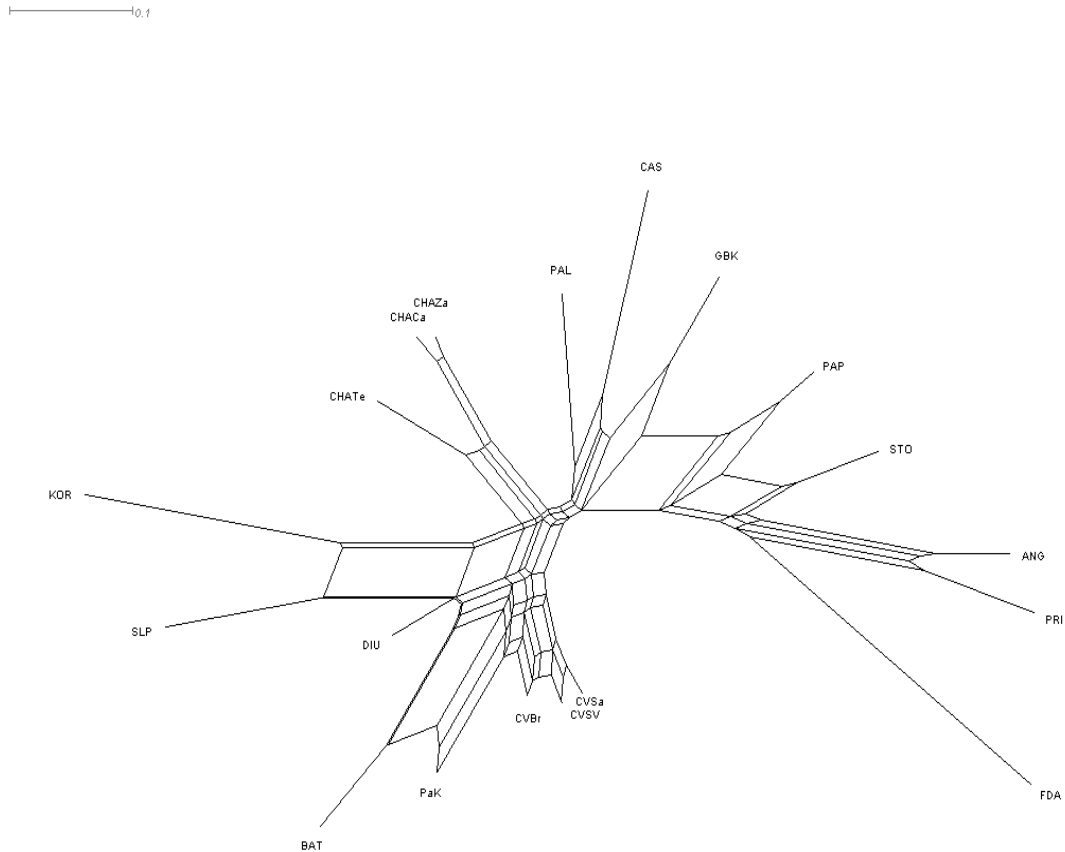


Figura 5. Rede de Proximidade – Neighbournet - dos 21 traços nominais dos crioulos ibéricos retirados do APiCS (Sippola & Tiny 2014).

Quando os 21 traços nominais são aplicados apenas aos 4 CGG, o resultado é a Figura 6 abaixo.

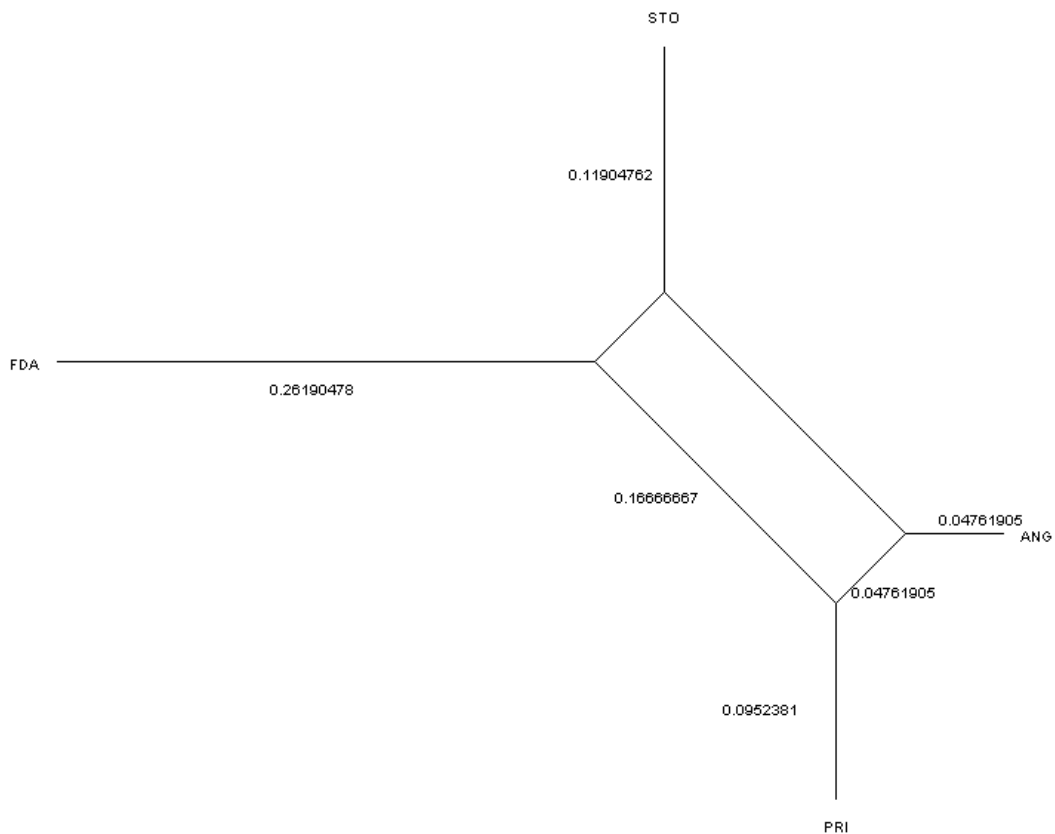


Figura 6. Rede de Proximidade – Neighbournet - dos 21 traços nominais dos CGG retirados do APiCS (Sippola & Tiny 2014)

Para análise dos dados, elaborámos a seguinte tabela relativa às distâncias a que as línguas se encontram entre si:

Pares	Distância
LI - ANG	0.1904
ST – ANG	0.3333
ST – LI	0.3808
ST – FA	0.4285
LI - FA	0.5237
ANG - FA	0.5237

Tabela 10. Distância numérica entre os CGG - traços nominais.

De acordo com a análise das Figuras 5 e 6, verifica-se que o comportamento dos CGG é relativamente constante. Na Figura 5, no que diz respeito aos CGG, o ANG e o LI (PRI) apresentam-se no mesmo nó, revelando considerável grau de proximidade. O ST (STO) e o FA (FDA) ocorrem em nós diferentes e as suas extremidades apresentam-se bastante distantes. Da análise das Figuras verifica-se ainda uma tendência de bipartição da rede segundo a qual os 4 CGG e o Papiamento parecem estar a ser encostados a uma extremidade da rede.

A Figura 6 vem ao encontro da tendência revelada na Figura 5. Pode-se verificar que o ANG e o LI se encontram mais próximos entre si (a 0.1904 de distância) e mais próximos do centro de informação partilhada. Por outro lado, e na extremidade oposta, o FA e o ST, embora partilhem o mesmo vértice, encontram-se bastante distantes, estando o ST mais próximo do ANG e do LI do que do FA. O FA, tal como na Figura 5, é o que está mais à margem, numa posição mais periférica em relação aos restantes.

A dimensão do centro de dados partilhados permite fazer a leitura inversa em relação à partilha de dados. Com base na Figura 6 é possível inferir que a proporção de informação comum deverá ser reduzida, pois o centro apresenta uma larga dimensão.

3.2. Dados Sintáticos

O objeto de estudo deste trabalho são os quatro CGG, que se encontram naturalmente incluídos no APiCS e apresentam por isso uma quantidade considerável de dados já trabalhados. O APiCS apresenta um número considerável de combinações nas diferentes

áreas da gramática a que se dedica. O APiCS pretende dar conta de traços, até certo ponto, generalizáveis nas 76 línguas, pelo que decidimos, para este estudo, analisar apenas os traços de natureza sintática constantes no APiCS, como já referido anteriormente. Tendo em conta que as línguas têm comportamentos diferenciados, extraímos os traços sintáticos existentes no APiCS e acrescentámos novos traços com características mais específicas e direccionadas para as línguas em estudo, totalizando um conjunto de 85 traços. Deste modo, com base nas respostas dadas pelos diferentes autores, cada um dos traços foi analisado e devidamente classificado de forma a ser introduzido no programa das redes filogenéticas. Veja-se o exemplo abaixo de um traço do APiCS e da respetiva classificação proposta. Nesta tabela, o número que se encontra no final da frase corresponde ao número da classificação do APiCS.

	Traços	ST	LI	ANG	FA
7	qual a ordem entre o demonstrativo e o nome [5]				
	demonstrativo-nome				
	nome-demonstrativo	+	+	+	+

Tabela 11. Traço relativo à ordem de palavras

A leitura que se faz desta tabela é a de que, nos quatro crioulos, a ordem entre os demonstrativos e o nome é a mesma, o que pode ser certificado com as respostas dos autores, ilustradas nos exemplos a seguir.

1) *Vinpema se.* (ST, Hagemeijer 2013, APiCS online)

Vinho palma DEM

‘Este vinho de palma / O vinho de palma em questão’

(2) *Ine ningê ixila, ine rupê.* (LI, Maurer 2013, APiCS online)

PL pessoas DEM 3PL brancas

‘Aqueles pessoas são brancas’

- (3) *Turu kwa e ma alê e thêka fa [...]* (ANG, Maurer 2013, APiCS online)
 Todas coisas DEM REL rei DEM PROG dizer [...]
 ‘Todas estas coisas que o rei estava a dizer [...]’

- (4) *Xadyise.* (FA, Post 2013, APiCS online)
 casa-DEM
 ‘Esta casa’

No entanto, em certos casos, uma língua pode usar várias estratégias diferentes, consoante o contexto e consoante os falantes que a reproduzem. Uma língua pode ter, portanto, para um mesmo traço dois ou mais valores divergentes. Tomemos como exemplo o traço 11 da tabela abaixo.

	Traços	ST	LI	ANG	FA
65	As construções reflexivas são expressas através de: [87]				
	Expressões implícitas	+			+
	Pronomes reflexivos com corpo ou parte do corpo	+	+	+	+
	Pronomes reflexivos compostos com intensificador		+		

Tabela 12. Traço relativo às construções reflexivas.

Através da Tabela 11, é possível verificar que, para o mesmo traço, algumas das línguas em estudo apresentam mais do que um valor canónico, sendo que, por um lado, o ST e o FA apresentam valores coincidentes entre si e, por outro, o LI e o ANG apresentam valores divergentes entre si e com as restantes. Esta possibilidade aplica-se a uma língua que admite diferentes estratégias em diferentes contextos, como se pode verificar nos exemplos seguintes:

- (5) *N ga ba kenta.* (ST, Hagemeijer 2013, APiCS online)
 1SG TMA ir aquecer
 ‘Eu vou-me aquecer’

(6) *N ga da ubwê mu ku faka.* (ST, Hagemeijer 2013, APiCS online)

1SG TMA dar corpo POSS1SG PREP faca

‘Vou-me esfaquear’

(7) *M mendu igbê me mesu.* (LI, Maurer 2013, APiCS online)

1SG medo corpo POSS1SG mesmo

‘Tenho medo de mim mesmo’

(8) *M mendu ami mesu.* (LI, Maurer 2013, APiCS online)

1SG medo 1SG mesmo

‘Tenho medo de mim mesmo’

(9) *Ê mata ôngê rê.* (ANG, Maurer 2013, APiCS online)

3SG matar corpo dele

‘Ele matou-se’

(10) *Dyia beza poxodulu bi sxa laba awa poto sai.* (FA, Post 2013, APiCS online)

Dia já pessoas ir TMA lavar água lago DEM

‘Antigamente as pessoas costumavam lavar-se nesse lago’

Em casos como os exemplos acima, em que um único traço tem diferentes valores canônicos, há necessidade de adaptação aquando da transposição para a tabela a ser introduzida no programa das redes filogenéticas. As tabelas a introduzir no programa são numéricas, pois este faz os cálculos apenas através da leitura que faz da numeração, coincidente ou divergente. Nestes casos, há a necessidade de optar por um valor numérico único que abranja os diferentes valores recolhidos para cada traço. Desta forma, a tabela abaixo (Tabela 13) reflete a classificação atribuída ao traço 65 (cf. Tabela 12). Note-se que ao ST e ao FA, que apresentam valores coincidentes na tabela anterior, se atribui a mesma numeração nesta tabela. Por sua vez, o LI e o ANG, tal como na tabela anterior, apresentam numeração divergente entre si e das restantes línguas.

	65
	As construções reflexivas são expressas através de: [87]
ST	1
LI	2
ANG	3
FA	1

Tabela 13. Tabela numérica relativa ao programa *SplitsTree4* (Traço 65)

Há ainda outros casos a considerar no processo de análise e classificação dos dados. A ordem entre os adjetivos exemplifica outra questão a considerar, na medida em que a ordem entre o nome e o adjetivo, nas quatro línguas, apresenta um valor único. No entanto, existem contextos em é permitida a inversão desta ordem; vejam-se os exemplos seguintes:

- (11) *Ũa soya glavi ku mwala glavi* (ST, Hagemeyer 2013, APiCS online)
 ART história bonita PREP mulher bonita
 ‘Uma história bonita com uma mulher bonita’

- (12) *Bon afe vs ma afe; bon ngê vs ma ngê.* (ST, Hagemeyer 2013, APiCS online)
 boa fé vs má fé; boa pessoa vs má pessoa
 ‘Boa / má fé; boa / má pessoa’

- (13) *Lonswe baanku.* (LI, Maurer 2013, APiCS online)
 Lençol Branco
 ‘Um lençol branco’

- (14) *N rêsbê ma nutixya.* (LI, Maurer 2013, APiCS online)
 1SG receber má notícia
 ‘Recebi más notícias’

(15) *M bê ãa buru ngai.* (ANG, Maurer 2013, APiCS online)
 1sg ver art pedra grande
 ‘Eu vi uma pedra grande’

(16) *Ũa bwa ngê.* (ANG, Maurer 2013, APiCS online)
 ART good person
 ‘Uma boa pessoa’

(17) *Xalafa vedyi.* (FA, Post 2013, APiCS online)
 Garrafa verde
 ‘A garrafa verde’

(18) *Panu véiu ~ véiu panu > vé panu* (FA, Post 2013, APiCS online)
 Vestido velho ~ velho vestido > velho vestido
 ‘Vestido velho’

Tendo em conta estas especificidades, tornou-se imperativo encontrar uma estratégia que permitisse uma leitura clara e objetiva da tabela e dos valores atribuídos, evitando leituras ambíguas. Optamos, nestes casos específicos, pela marcação do valor relativo à ordem canónica da língua quando esta apresenta um padrão canónico. Por canónico entendemos a opção que apresente maior percentagem de ocorrência. Esta percentagem foi atribuída por cada colaborador responsável pela classificação de uma ou mais línguas do seu domínio e indicado no APiCS online. Assim, os dados representados em (11)-(18) foram classificados da seguinte forma:

	Traços	ST	LI	AN G	FA
4	qual a ordem canónica entre os adjetivos e o nome [3]				
	nome-adjetivo	+	+	+	+
	adjetivo-nome				

Tabela 13. Traço relativo a ordem de sintagmas

Num caso específico, houve necessidade de alterar uma resposta do APiCS, em virtude da obtenção de novos dados para o FA. Veja-se a Tabela 14 abaixo, em que o valor do FA sofreu alteração, sendo que o valor que está a vermelho foi o valor atribuído pelo APiCS e o novo valor encontra-se acima.

14	qual a posição que ocupa o marcador de plural em relação ao nome [23]				
	pré- nominal	1	1	1	1
	pós-nominal				
	a língua não possui um marcador nominal de plural				1

Tabela 14. Traço relativo à ordem do marcador de plural em relação ao nome.

Atente-se no exemplo seguinte produzido pelos falantes:

- (19) *Nan khamada mu* (FA, corpus do Fa d’Ambô, CLUL)
 PL camarada POSS
 ‘Os meus camaradas’

Após a análise dos dados, introduziu-se o ficheiro no programa das redes filogenéticas e obtiveram-se as seguintes redes:

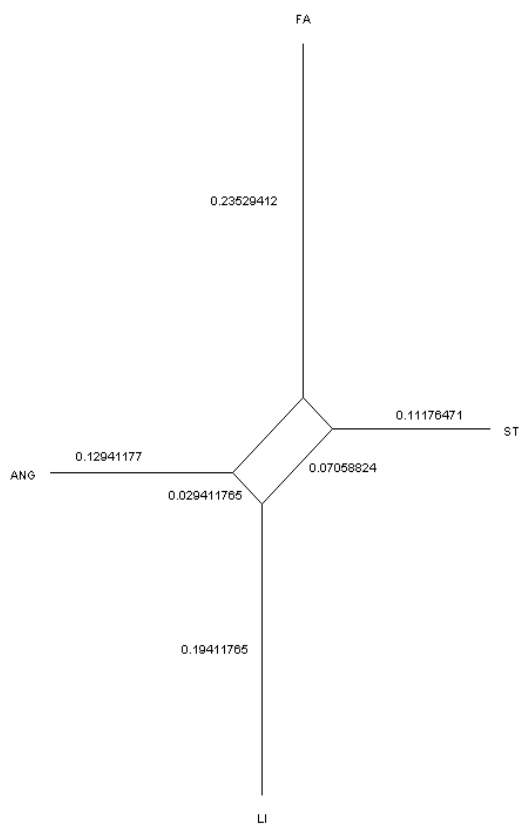


Figura 7. Lista de traços sintáticos aplicada à Rede de Proximidade (Neighbournet)

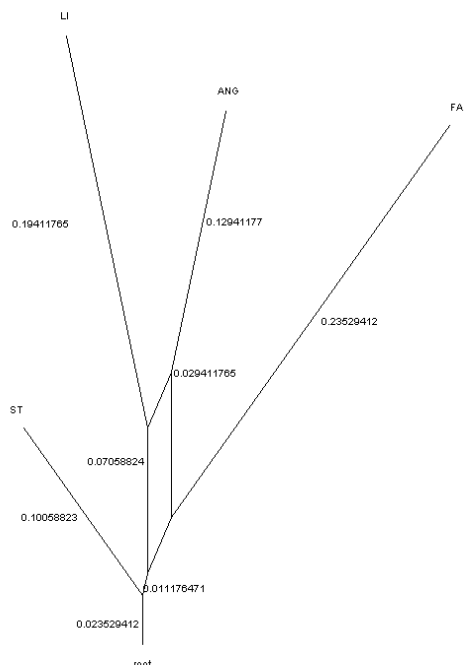


Figura 8. Lista de traços sintáticos aplicada à Rede com Raiz (*Rooted Network*)

As distâncias entre cada par de línguas a este respeito são indicadas na Tabela 15.

Pares	Distância
ST – ANG	0.3410
LI – ANG	0.3529
ST – LI	0.3763
ST – FA	0.3763
ANG - FA	0.4351
LI - FA	0.5292

Tabela 15. Distância numérica entre os CGG – lista de traços sintáticos

3.2.1. Análise

Ao analisar a Figura 7, verifica-se que o centro de traços partilhados apresenta uma dimensão considerável o que significará que existe um elevado número de traços não

partilhados entre os quatro CGG. Esta leitura vai ao encontro do que se pode verificar na correspondente rede com raiz. Nesta, a distância de todas as línguas à raiz é consideravelmente superior quando comparada com as restantes redes de raiz produzidas no decurso deste estudo. O ST é o que se apresenta mais próximo do centro, imediatamente seguido do ANG, que apresenta uma distância ao centro pouco superior ao ST. O FA e o LI encontram-se em extremos opostos da Figura 7, o que levará a concluir que o LI e o FA são as línguas mais distantes entre si e que menos traços partilham. No entanto, de acordo com o quadro das distâncias, podemos verificar que estas duas línguas se encontram precisamente à mesma distância do ST.

O FA, apesar de ser o que se apresenta mais distante do centro, encontra-se numa posição próxima do ST, na medida em que se encontram ambos no mesmo vértice da rede. Esta proximidade é reforçada quando se analisa a rede com raiz, pois esta permite verificar que o nó onde o FA ramifica se encontra mais próximo do nó onde o ST ramifica. Por outro lado, a proximidade entre o LI e o ANG, tanto a nível da distância entre si como o facto de se encontrarem no mesmo vértice na rede, é reforçada pela proximidade do nó em que ambos ramificam na rede com raiz.

3.3. Conclusão

Ao compararmos os dados das Figuras 6 e 7 relativas à sintaxe nominal e ao levantamento dos traços sintáticos em geral, verifica-se que há uma clara tendência para uma bipartição entre o ANG e o LI, por um lado, e o FA e o ST, por outro. Esta tendência é claramente verificável nas redes de proximidade 5, 6 e 7 que, nos três casos, apresentam o LI e o ANG no mesmo vértice da raiz. Verifica-se também que os dados partilhados em ambos os casos é relativamente baixo, tendo em conta que o tamanho do centro de partilha é amplo em ambas as redes, levando a um espraio da figura e a uma distância relativamente ampla entre os extremos.

A análise das figuras permite ainda verificar que o FA é o que se apresenta consistentemente mais afastado do centro, apoiando a ideia de que o FA permaneceu em isolamento em relação às restantes línguas, tendo seguido o seu percurso de forma

independente. No entanto, verifica-se que entre o ST e o FA há uma relação constante de proximidade, o que nos dirá que estas línguas divergem no mesmo sentido.

Os dados relativos ao SN e aos traços sintáticos no geral apresentam relativa complementaridade, o que permite verificar uma clara tendência.

4. Conclusões finais

Tal como a generalidade dos estudiosos que trabalham sobre os CGG, assumimos como ponto de partida deste estudo que os CGG constituem uma família linguística, isto é, os quatro CGG descendem historicamente de um antepassado comum: o proto-crioulo do Golfo da Guiné. A partilha de um grande número de traços lexicais e sintáticos (cf. Anexos) corrobora esta hipótese.

Para este estudo comparado, recorremos ao uso de redes filogenéticas. Este método foi concebido inicialmente para estudos na área da biologia, mas tem vindo a popularizar-se na área da linguística. Neste sentido, de forma a identificar cognatos, aplicámos a lista das 200 palavras de léxico básico de Swadesh aos quatro CGG. Preparámos ainda uma lista de léxico funcional no sentido de perceber o comportamento deste subconjunto do léxico, especialmente em comparação com o léxico básico. Após classificarmos os dados, estes foram introduzidos no programa das redes filogenéticas, o *SplitsTree4*. Procedemos igualmente ao levantamento dos traços sintáticos que constam do APiCS, que contém 130 traços distintos, e, de forma a direccionar os traços para as especificidades destes crioulos, introduzimos um conjunto de novos traços sintáticos. O resultado foi uma lista de 85 traços. Depois de classificados, estes foram introduzidos no programa de redes filogenéticas. A partir das redes que foram geradas, foi-nos possível avançar com algumas considerações.

Através das distâncias observadas nas redes correspondentes ao léxico básico e funcional, constatamos que o centro da Figura 1, que corresponde aos traços partilhados, é bastante reduzido e uniforme, o que nos permite inferir que os CGG partilham uma quantidade superior de léxico básico do que de léxico funcional.

O ST tem sido considerado a continuação do PCCG no tempo e no espaço. Neste sentido, o facto de os dados evidenciarem que o ST é o que se apresenta consistentemente mais próximo do centro de dados partilhados vem conceder ao ST um papel de centralidade em relação aos restantes. Isto corrobora a visão de que o PCCG se terá formado na ilha de São Tomé, no mesmo espaço ocupado hoje pelo ST.

Em relação ao léxico, o LI é o que se encontra claramente mais próximo do ST em ambas as redes, sendo que, na rede da Figura 1, estas duas línguas se apresentam praticamente à mesma distância. Estes dados confirmam o que foi defendido por Ferraz (1979), que considerava que havia uma maior partilha de léxico entre o ST e o LI. Estes resultados indicam uma maior proximidade entre o ST e o LI, possivelmente devido a algum contato histórico continuado.

Analisando apenas os dados sintáticos, é possível verificar que há clara tendência de bipartição que junta o ANG e o LI, por um lado, e o FA e o ST, por outro. No entanto, o ST, embora se encontre no mesmo vértice da rede que o FA, encontra-se a uma distância mais próxima do ANG e do LI, respetivamente, do que do FA. É possível verificar ainda que o FA se encontra numa posição significativamente mais exterior e distante dos restantes crioulos. Estima-se que a ilha de Ano Bom foi povoada em meados do século XVI e que a população que terá chegado a esta ilha nos séculos XVI e XVII já teria o proto-crioulo como língua nativa ou dominante. Como o isolamento do LI e do ANG parece ter sido anterior, o maior grau de divergência do FA com os demais crioulos poderá ser atribuído ao grande isolamento em que esta ilha sempre se encontrou (Hagemeijer & Zamora, no prelo) depois do povoamento definitivo.

Tanto a nível do léxico básico como a nível do léxico funcional, o ANG é o que se apresenta mais distante das restantes línguas e do centro de traços partilhados, o que não se verifica em relação à sintaxe. Relacionamos este dado com o impacto específico das línguas Bantu neste crioulo. Estas línguas terão sido trazidas na fase de plantação, em que os escravos passaram a ser resgatados na região do Congo e de Angola, em detrimento da região do delta do Níger. Por razões que possam ter a ver com a composição etno-linguística dos escravos que fugiam das plantações, o ANG terá sofrido um processo de relexificação, tendo o léxico crioulo original sido substituído por léxico Bantu, em particular do Kimbundu (Maurer 1992, Lorenzino 1998). Explica-se, assim, a maior distância do ANG em relação aos demais CGG nas redes lexicais apresentadas. A maior proximidade sintática entre o ST e o ANG poderá ser atribuída à existência de contacto entre estas duas línguas faladas na mesma ilha. .

Conclui-se que a utilização de redes filogenéticas constitui uma forma mais objetiva de calcular e compreender as distâncias entre línguas e, no caso específico da

família dos CGG, de confrontar os resultados com as hipóteses que têm vindo a ser defendidas sobre a sua origem e desenvolvimento. Consideramos que as conclusões obtidas neste estudo indicam tendências que possam ser fortalecidas ao alargar e refinar ainda mais o leque de traços comparados.

Referências Bibliográficas

- Araújo, Gabriel & Hagemeijer, Tjerk. 2013. *Dicionário livre do santome-português*. São Paulo: Hedra.
- Bakker, Peter; Daval-Markussen, Aymeric; Parkvall, Mikael; Plag, Ingo. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26(1): 5-42.
- Caldeira, Arlindo. 2004. Rebelião e outras formas de resistência à escravidão nas ilhas do Golfo da Guiné (séculos XVI-XVIII). *Studia Africana* 7. 101-136.
- Caldeira, Arlindo. 2006. Uma ilha quase desconhecida. Notas para a história de Ano Bom. *Studia Africana – Revista Interuniversitária d’Estudis Africans* 17. 99-109.
- Decreto de lei nº19/2013 a 14 de agosto de 2013, no Diário da República de São Tomé e Príncipe
- Ferraz, Luiz Ivens. 1974. A linguistic appraisal of Angolar. *In Memoriam Antonio Jorge Dias*, Vol. 2, 177-186. Lisbon: Instituto de Alta Cultura/Junta de Investigações Científicas do Ultramar.
- Ferraz, Luiz Ivens. 1976. The origin and development of four creoles in the Gulf of Guinea. *African Studies* 35(1). 33-38.
- Ferraz, L. 1979. *The creole of São Tomé*. Johannesburg: Witwatersrand University Press.
- Ferraz, Luiz Ivens. 1987a. The liquid in the Gulf of Guinea Creoles. *African Studies* 46(2): 287-95.

- Ferraz, L. 1987b. *Portuguese creoles of West Africa and Asia*. In Gilbert, Glen (ed.), *Pidgins and creole languages: Essays in memory of John E. Reinecke*, 337-60. Honolulu: University of Hawaii Press.
- Günther, Wilfried. 1973. *Das Portugiesische Kreolisch der ilha do Príncipe*. Marburg an der Lahn: Marburger Studien zur Afrika- und Asienkunde.
- Hagemeijer, Tjerk. 2009. As línguas de S. Tomé e Príncipe. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola*, 1:1, 1-27.
- Hagemeijer, Tjerk. 2011. The Gulf of Guinea creoles: genetic and typological relations. *Journal of Pidgin and Creole Languages*. 26:1, 111-154.
- Hagemeijer, Tjerk & Armando Zamora (no prelo). Fa d'Ambô: Past and present. *International Journal of the Sociology of Language*.
- Huson, Daniel H. & David Bryant. 2006. *Application of phylogenetic networks in evolutionary studies*. Mol. Biol. Evol. 23 (2). 254 - 267
- Lorenzino, Gerardo. 1998. *The Angolar Creole Portuguese of São Tomé: Its grammar and sociolinguistic history*. Ph.D. dissertation, City University of New York.
- Matos, Raimundo José da Cunha. 1842. *Corographia histórica das ilhas de S. Thomé, Príncipe, Anno Bom e Fernando Pó*. Porto: Typographia da Revista.
- Maurer, Philippe. 1995. *L'Angolar: Un créole afro-portugais parlé à São Tomé*. Hamburg: Helmut Buske Verlag.
- Maurer, Philippe 1999. El verbo locativo poner en santomense, principense y angolar. In Klaus Zimmermann (ed.), *Lenguas criollas de base lexical española y portuguesa*, 89-100. Frankfurt/Madrid: Vervuert/ Iberoamericana.

- Maurer, Philippe. 2009. *Principense — Grammar, texts, and vocabulary of the Afro-Portuguese creole of the island of Príncipe*. London: Battlebridge Publications.
- McMahon, April & Robert McMahon. 2003. Finding families: Quantitative methods in language classification. *Transactions of the Philological Society* 101. 7–55.
- Michaelis, Susanne; Maurer, Philippe; Haspelmath, Martin; Huber, Magnus (eds.). 2013. *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Michaelis, Susanne; Maurer, Philippe; Haspelmath, Martin; Huber, Magnus (eds.). 2013. *The survey of pidgin and creole languages. Volume 2: Portuguese-based, Spanish-based, and French-based Languages*. Oxford: Oxford University Press
- Negreiros, Almada José. 1895. *Historia Ethnographica da ilha de S. Tomé*. Lisboa.
- Schang, Emmanuel. 2000. *L'émergence des créoles portugais du Golfe de Guinée*. Tese de Doutoramento, Universidade de Nancy 2. .
- Seibert, Gerhard. 2007. Angolares of São Tomé island. In Philip Havik & Malyn Newitt (eds.), *Creole societies in the Portuguese colonial empire*, 105-126. Bristol: Bristol University Press.
- Sippola, Eeva & Tiny, Abigail. 2014. Noun Phrases in Iberian creoles. Aarhus University, January 15-16, 2014. Time and Space in Linguistics: Interdisciplinary Computational Approaches & Cross-Creole Comparisons. (disponível em: <http://tasil2014.com/12sippolatiny-ppt.pdf>)
- Sousa, Celso Batista de. 1990. *S. Tomé e Príncipe. Do descobrimento aos meados do século XVI*. Dissertação de Mestrado, Universidade de Lisboa.

Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*, 21:2, 121 – 137.

Valkhoff, Marius. 1966. *Studies in Portuguese and creole*. Johannesburg: Witwatersrand University Press.

Zamora, Armando. 2010. *Gramática descriptiva del fa d'ambô*. Barcelona: CEIBA Ediciones.

Corpora consultados:

<http://alfclul.clul.ul.pt/CQPweb/Angolar>

<http://alfclul.clul.ul.pt/CQPweb/Fadambo/>

<http://alfclul.clul.ul.pt/CQPweb/Principense/>

<http://alfclul.clul.ul.pt/CQPweb/Santome/>

<http://apics-online.info/>

ANEXOS

ANEXO I - Lista de 200 palavras de Swadesh: Origem do vocabulário

Legenda							
1=Português							
2=Africano							
3=Outro/desconhecido							
Inglês		Santome		Lung'Ie		Angolar	Fa d'Ambô
all	1	tudu	1	tudu	1	turu	1 tudu
and / with	1	ku	1	ki	1	ki	1 ku
animal	1	bisu	1	bisu	1	bithu	1 limaya
ash	1	xinza	1	ixize	1	xindja	1 xinza, sinza
back	1	tlaxi	1	taaxi	2	nthusi	1 lomba
bad	3	bluku	3	buuku	3	buuku	3 buuku
bark (of tree)	1	kaxka	2	ukpaka	1	kaxka	1 khokha
beard	1	beba	1	bweba	2	fantxi	1 balba
because	3	punda	3	pidi	3	punda	1 pokê
belly	1	bega	1	bwega	1	bega	1 beega
big	1	glandji	1	gaani	1	ngai	1 gaanyi
bird	1	bisu	1	pasu	3	situ	1 patu
black	1	pletu	1	peetu	1	peetu	1 peetu
blood	1	sangi	1	isengi	1	thangi	1 sangi
bone	1	oso	1	osu	1	otho	1 oso
child (a youth)	1	mina	1	minu	1	nna	1 mina
cloud	1	nôvi	1	nôvi	3	mbosi	1 nôvi
cold (as in weather)	1	fyô	1	fyô	1	fiô	1 fiiu
cow	1	bwê	1	ubwê	1	buê	1 ôbôyô
day (daytime)	1	dja	1	dya	1	dia	1 dja
death	1	motxi	1	motxi	1	motxi	1 mootxi
dirty	1	suzu	1	suzu	1	thudhu	1 sunzu
dog	1	kasô	1	kasô	1	kathô	1 khôsôlu
dry	1	seku	1	seku	1	theku	1 seku
dust	1	opo	1	opo	1	opo	1 opo
ear	1	olha	1	urya	1	oria~odia	1 ôlea
earth (as in soil)	1	tela	1	ite	3	mavu	1 tele
egg	1	ovu	1	ovu	1	ôvu	1 ovu
eye	1	wê	1	uwê	1	wê	1 oyu
far	1	lonji	1	lonji	1	londji	1 lonji
fat	1	banya, banha	1	banha	1	banya	1 gudu
father	1	pe	1	pwe	2	tata	1 pe
feather (rather not down)	1	pena	1	pene	1	pena	1 pena
fingernail	1	inhe	1	umunya	1	inyê	1 inha

fire	1	fôgô	1	ufôgô	1	fôgô	1	fôgu
fish	1	pixi	1	pêxi	3	kikiê	1	pixi
five	1	xinku	1	xinku	2	tano	1	xinku
flower	1	foli	1	foli	1	foori	1	fool
foot	1	ope	1	ope	1	ope	1	ôpe
forest	2	ôbô	2	ôvyô	2	ôbô	2	ôgô
four	1	kwatlu	1	kwatu	2	kuana	1	khatul
full	1	xa	1	xyadu	1	siaru	1	xia
good	1	bon	1	bon	1	bo	1	bon
green	1	vêdê	1	vêdê	1	dhulu(ru)	1	vêyi
ground	1	son	1	usan	1	thon	1	san
hand	1	mon	1	uman	1	mon	1	oman
he	1	ê	1	ê	1	ê	1	ê
head	1	kabesa	1	kabese	3	ntê	1	khaasa
heart	1	kloson	1	kosan	1	kotho	1	kusan
here	1	nai	1	ni	3	inge	1	lhai
horn	1	xifli	1	xifi	1	koneta	3	khôkhô
house	1	ke	1	kaxi	1	khadji	1	kai
I	3	n	3	n	3	m	3	mu
if	1	xi	1	xi	1	si	1	si
impersonal pronoun	2	a	2	a	2	a	2	a
in, at	1	ni	1	na	1	ni	1	n
inside	1	nglentu	1	udentu	1	lêtu	1	dantulu
knee	1	zê	1	ôzê	2	mpuna	1	ôzôyô
knife	1	faka	1	ufaka	1	faka	1	fakha
leaf	1	fya	1	ufya	1	fia	1	fa
leg	1	ope	1	ope	1	ope	1	ôpe
liver	1	figadu	1	figadu	1	figaru	1	fugudu
long	1	lôngô	1	lôngô	1	lôngô	1	lôngô
louse	2	idu	2	idu	2	iru	2	idu
man (adult male)	1	ome	1	omi	1	ome	1	omi
man (human being)	1	ngê	1	ningê	1	ningê	1	ngê
many	1	montxi	1	montxi	1	motxi	1	montxi
meat (as in flesh)	1	kani	1	kani	1	kani, situ	1	khanyi
moon	1	nwa	1	unwan	3	mbêzi	1	ônuje
mother	1	men	1	mwin	1	mama	1	men
mountain	2	ôkê	2	ôkyê	2	lôndê	2	iki
mouth	1	boka	1	ubuka	1	boka	1	boXo
name	1	nomi	1	nomi	1	nomi	1	nomi
navel	1	binku	1	bigu	1	mbiku	3	okho
near	1	petu	1	petu	1	petu	1	peetu
neck	3	klonklo	3	ôtô	3	singô	3	godxi
new	1	novu	1	novu	1	novu	1	novu
night	1	nôtxi	1	unôtxi	3	nthuku	1	nôtxi
nose	1	lixi	1	irixi	1	lisi	1	lix
not (final negation marker)	3	fa	3	fa	3	wa	3	f
old	1	ve	1	ve	3	kuka	1	ve

on top of	1	liba	1	ixima	1	riba	1	liba
one	1	ũa	1	ũa	1	ũa	1	wan
other	1	ôtlô	1	ôtô	1	ôtô	1	utulu
poor	1	pobli	1	pobê	1	pobi	1	pôvili
red	1	vlêmê	1	vêmê	2	bôbô	2	bôbô
road	1	kamya	1	kumin	3	mondja	1	xama
root	1	leji	1	reji	3	ndatxi	1	lezê
rope	1	kodo	1	kwaa	1	kodo	1	khodo
round	1	londondô	1	rodondo	3	kondogaru	1	lodondo
salt	1	salu	1	usalu	1	thalu	1	salu
sand	1	alya	1	arya	3	nthêkê	1	alea
sea (as in ocean)	1	omali	1	umwen	2	mionga	1	omal
seed	2	ukwê	2	ikpe	2	ikwê~inkwê	2	iku
serve, work for	1	xtlivi	1	xivi	1	sivi	1	xiivi
short	1	kutu	1	kutu	1	kutu	1	kuutu
skin (of a person)	1	pele	1	pele	1	pele	1	pele
sky	1	ose	1	ose	1	onthe	1	ôsê
smoke	2	igligô	2	igigu	2	irigô	2	igôgu
snake	1	koblo	1	kôbô	3	kiongo	1	khôbôlô
soul	1	alima	1	alima	1	alima	1	alma
spit	1	kupi	1	kupi	1	kupi	1	tabokho
star	1	stlela	1	teela	2	têtêmbu	1	ntela
stick (of wood)	1	po	1	upa	1	po	1	opa
stone	2	budu	2	ubudu	2	buru	2	budu
stone of a fruit	1	klôsô	1	klôôsu	1	kôôthu	1	xôôsu
straight	1	dlêtu	1	dêêtu	1	deeto	1	dêêtu
sun	1	solo	1	isolu~usolu	1	tholo	1	sôlô
tail	1	labu	1	urabu	3	inkila	1	labu
tall	1	longô	1	longô	1	longô	1	longô
there	1	ala	1	la	1	nha	1	ala
they	2	inen	2	ine	2	ene	2	êney
thick (fat)	1	gôdô	1	gôdô	3	netaru	1	gôdô
thin (slim)	1	mlagu	1	mangu	1	mangu	1	mungulu
thou (sg.)	1	bô	1	txi	1	bô	1	bo
three	1	tlêxi	1	têêxi	1	têêsi	1	têêxi
to ask	1	punta	1	punta	1	puta	1	pologunta
to bark	1	ladla	3	kupa	3	budja	1	ladala
to bite	1	môdê	1	môdê	1	môrê	1	munda
to burn	1	kema	1	kema	1	kema	1	khama
to bury	1	ntela	1	kava	1	tia	1	ntela
to come	1	bi	1	vika	1	bi	1	bi
to count	1	konta	1	konta	1	konta	1	khonta
to cultivate	1	ximya	1	ximya	1	ximya	1	ximya
to cut	1	kota	1	kota	1	kota	1	khoota
to die	1	môlê	1	mwê	1	muê	1	môlê
to dig	1	kobo	1	kava	3	vina	1	khobo
to drink	1	bêbê	1	bêbê	1	bêbê	1	bebê
to eat	1	kume	1	kume	1	mme	1	kumu

to fall (as in drop)	1	kyê	1	kyê	1	kiê	1	kêê
to fear	1	mendu	1	mendu	1	mendu	1	mendu
to fight	1	luta	1	luta	3	vugu	1	luta
to float	1	boya	1	boya	1	boya	1	boya
to fly	1	vwa	1	vwa	3	pupuka	1	va
to follow, to accompany	2	lêlê	2	lêlê	2	lêlê	2	lêlê
to give	1	da	1	da	1	da	1	da
to give birth	1	pali	1	pwe	1	pari	1	pali
to hear	1	têndê	1	têndê	1	êndê	1	tindi
to hunt (transitive)	1	kasa	1	kasa	1	katha	1	khasa
to kill	1	mata	1	mata	1	mata	1	mata
to know (a fact)	1	sêbê	1	sêbê	3	eta	1	sabê
to know a person	1	konsê	1	konsê	1	konthê	1	khonsê
to laugh	1	li	1	rin	1	di	1	li
to lift up	1	ligi	1	rêgê	1	rigi	1	ligi
to live	1	ta	1	ta	1	ta	1	ta
to play	1	floga	1	foga	1	foga	1	fuga
to pull	1	saya	1	saa	1	thaa	1	saa
to push	1	pinsa	1	pinsa	1	pintha	1	piza
to rain	1	sôbê	1	usuva	1	thuba	1	xiba
to request, to ask	1	pidji	1	pidi	1	pindji	1	pindji
to rub	1	flega	1	fega	1	fega	1	feega
to run	1	kôlê	1	kwê	1	kôlê	1	khôlê
to say	1	fla	1	fa	1	fa	1	fa
to scratch (an itch)	1	kosa	1	kosa	1	kotha	1	kusa
to see	1	bê	1	vê	1	bê	1	bê
to send	1	manda	1	manda	1	manda	1	manda
to sew	1	klôzê	1	kôzê	1	kôdhê	1	khôzê
to show	1	musa	1	musa	1	uthua	1	minsia
to sing	1	kanta	1	kanta	1	kata	1	khanta
to sit	1	tason	1	tusan	1	tatho	1	tusan
to sleep	1	dumini	1	dimi	1	dumi	1	djuni
to smell (sense odor)	1	sela	1	sera	1	thela	1	sela
to split (wood)	2	va	2	va	1	taya	1	lasa
to squeeze	1	peta	1	peta	1	peta	1	peeta
to stand up	1	lanta	1	lata	1	lata	1	laanta
to steal	1	futa	1	futa	1	futa	1	futa
to suck	1	supa	1	supa	1	thupa	1	supa
to swell	1	fuma	1	fuma	1	fuma	1	fuma
to swim	1	landa	1	landa	1	landa	1	laanta
to switch, exchange	1	tloka	1	toka	1	toka	1	tuuka
to think	1	pensa	1	pensa	1	pentha	1	pensa
to throw	1	zuga	1	zuga	1	dhuga	1	zuga
to turn (change direction)	1	bila	1	vya	1	bila	1	vla
to vomit	1	lansa	1	lansa	1	ngumita	1	luza
to walk	1	nda	1	nda	1	nda	1	nda
to want	1	mêsê	1	mêsê	1	mêthê	3	ngongô
to wash	1	laba	1	lava	1	laba	1	laba

to wet	1	monha	1	mwa	1	minha	1	mya
to whistle	1	supya	1	supya	1	thupya	1	supya
tongue	1	lungwa	1	lunge	1	lunga	1	lunge
tooth	1	dentxi	1	idintxi	1	detxi	1	dentxi
tortoise	1	tatalugwa	1	tetuga	1	têtêuga	1	tôtxiiga
tree	1	po	1	upa	1	po	1	opa
two	1	dôsu	1	dôsu	1	dôthu	1	dôsu
underneath	1	basu	1	ubasu	3	kosi	1	basu
urine	1	zawa	1	mize	1	dhawa	1	minza
war	1	gela	1	ge	1	ngela	1	gela
warm (as in weather)	1	kentxi	1	kyentxi	1	kêtxi	1	kentxi
water	1	awa	1	awa	1	awa	1	awa
we	1	non	1	non	1	no	1	non
what	1	kêkwa	1	kwa	1	kwai	1	kêkhwa
when (temporal)	1	ola	1	ora	1	ola	1	ola
white	1	blanku	1	baanku	3	zia	1	banku
who	1	kengê	1	ningê	1	ngêi	1	kengê
wind (as in breeze)	1	ventu	1	uventu	1	ventu	1	oventu
wing	1	aza	1	aza	1	adha	1	aza
woman	1	mwala	1	mye	2	mengai	1	myele
wound	1	flida	1	fiida	1	fiira	1	filida
year	1	anu	1	anu	1	anu	1	anu
you (pl.)	3	nansê	3	owo	3	êthê	3	naminsêji

ANEXO II - Lista de 200 palavras Swadesh: Cognatos – Versão para *SplitsTree4*

ENGLISH	Santome	Lung'le	Angolar	Fa d'Ambô	S T	L I	AN G	F A
all	tudu	tudu	turu	tudu	1	1	1	1
and / with	ku	ki	ki	ku	1	1	1	1
animal	bisu	bisu	bithu	limaya	1	1	1	2
ash	xinza	ixize	xindja	xinza, sinza	1	1	1	1
back	tlaxi	taaxi	nthusi	lomba	1	1	2	3
bad	bluku	buuku	buuku	buuku	1	1	1	1
bark (of tree)	kaxka	ukpaka	kaxka	khokha	1	2	1	3
beard	beba	bweba	fantxi	balba	1	1	2	1
because	punda	pidi	punda	pokê	1	2	1	3
belly	bega	bwega	bega	beega	1	1	1	1
big	glandji	gaani	ngai	ngaandji	1	1	1	1
bird	bisu	pasu	situ	patu	1	2	3	2
black	pletu	peetu	peetu	peetu	1	1	1	1
blood	sangi	isengi	thangi	sangi	1	1	1	1
bone	oso	osu	otho	oso	1	1	1	1
child (a youth)	mina	minu	nna	mina	1	2	1	1
cloud	nôvi	nôvi	mbosi	nôvi	1	1	2	1
cold (as in weather)	fyô	fyô	fiô	fiiu	1	1	1	1
cow	bwê	ubwê	buê	ôbôyô	1	1	1	1
day (daytime)	dja	dya	dia	dja	1	1	1	1
death	motxi	motxi	motxi	mootxi	1	1	1	1
dirty	suzu	suzu	thudhu	sunzu	1	1	1	1
dog	kasô	kasô	kathô	khasôl	1	1	1	1
dry	seku	seku	theku	seku	1	1	1	1
dust	opo	opo	opo	opo	1	1	1	1
ear	olha	urya	oria	olea	1	1	1	1
earth (as in soil)	tela	ite	mavu	tele	1	1	2	1
egg	ovu	ovu	ôvu	ovu	1	1	1	1
eye	wê	uwê	wê	ôyô	1	1	1	2
far	lôndji	lônji	lôndji	lôndji	1	1	1	1
fat (noun)	banha	banha	banha	gudu	1	1	1	2
father	pe	pwe	tata	pe	1	1	2	1
feather (rather not down)	pena	pene	pena	pena	1	1	1	1
finger nail	inhe	umunha	inhe	inha	1	2	1	1
fire	fôgô	ufôgô	fôgô	fôgô	1	1	1	1
fish	pixi	pêxi	kikiê	pixi	1	1	2	1
five	xinku	xinku	tano	xinku	1	1	2	1
flower	floli	foli	foori	fool	1	1	1	1
foot	ope	ope	ope	ôpe	1	1	1	1
forest	ôbô	ôvyô	ôbô	ôgô	1	1	1	1
four	kwatlu	kwatu	kuana	khatul	1	1	2	1
full	xa	xya	sia	xia	1	1	1	1

good	bôn	bôn	bo	bôn	1	1	1	1
green	vêdê	vêdê	dhulu(ru)	vêêdji	1	1	2	1
ground	son	usan	thon	san	1	1	1	1
hand	mon	uman	mon	ôman	1	1	1	1
he	ê	ê	ê	ê	1	1	1	1
head	kabêsa	kabese	ntê	khaasa	1	1	2	1
heart	kloson	kosan	kotho	kusan	1	1	1	1
here	nai	ni	inge	yai	1	2	3	1
horn	xifli	xifi	koneta	khônkhô	1	1	2	3
house	ke	kaxi	kai	khai	1	1	1	1
human being, person	ngê	ningê	ngê	ngê	1	1	1	1
I	n	n	m	mu	1	1	1	1
if	xi	xi	si	si	1	1	1	1
impersonal pronoun	a	a	a	a	1	1	1	1
in, at	ni	na	ni	en	1	2	1	3
inside	nglêntu	udêntu	lêtu	dantulu	1	1	1	1
knee	zê	ôzê	mpuna	ôzôyô	1	1	2	1
knife	faka	ufaka	faka	fakha	1	1	1	1
leaf	fya	ufya	fia	fa	1	1	1	1
leg	ope	ope	ope	ope	1	1	1	1
liver	figadu	figadu	figaru	fugudu	1	1	1	1
long	lôngô	lôngô	lôngô	lôngô	1	1	1	1
louse	idu	idu	iru	idu	1	1	1	1
man (adult male)	ome	omi	ome	ome	1	1	1	1
man (human being)	ngê	ningê	ningê	ngê	1	1	1	1
many	montxi	montxi	motxi	montxi	1	1	1	1
meat (as in flesh)	kani	kani	kani, situ	khaanyi	1	1	1	1
moon	nwa	unwan	mbêzi	ônwan	1	1	2	1
mother	men	mwin	mama	men	1	1	2	1
mountain	ôkê	ôkyê	lôndê	iki	1	1	1	1
mouth	boka	ubuka	boka	bôkha	1	1	1	1
name	nomi	nomi	nomi	nomi	1	1	1	1
navel	binku	bigu	mbiku	okho	1	1	1	2
near	petu	petu	petu	peetu	1	1	1	1
neck	klonklo	ôtô	singô	khôkhôndj ô	1	2	3	4
new	novu	novu	novu	novu	1	1	1	1
night	nôtxi	unôtxi	nthuku	nôtxi	1	1	2	1
nose	lixi	irixi	lisi	lixi	1	1	1	1
not (final negation marker)	fa	fa	wa	f	1	1	1	1
old	ve	ve	kuka	ve	1	1	2	1
on top of	liba	ixima	riba	liba	1	2	1	1
one	ũa	ũa	ũa	wan	1	1	1	1
other	ôtlô	ôtô	ôtô	utulu	1	1	1	1
poor	pobli	pobê	pobi	pôvili	1	1	1	1
red	vlêmê	vêmê	bôbô	bôbô	1	1	2	2
road	kamya	kumin	mondja	khama	1	1	2	1
root	leji	reji	ndatxi	lezê	1	1	2	1

rope	kodo	kwaa	kodo	khodo	1	2	1	1
round	londondo	rodondo	kondogaru	lodonda	1	1	2	1
salt	salu	usalu	thalu	salu	1	1	1	1
sand	alya	arya	nthêkê	alea	1	1	2	1
sea (as in ocean)	omali	umwen	mionga	omali	1	2	3	1
seed	ukwê	ikpe	ikwê~inkwê	iku	1	1	1	1
serve, work for	xtlivi	xivi	sivi	xiivi	1	1	1	1
short	kutu	kutu	kutu	kuutu	1	1	1	1
skin (of a person)	pele	pele	pele	pele	1	1	1	1
sky	ose	ose	onthe	ôsê	1	1	1	1
smoke	igligô	igigu	irigô	igôgô	1	1	1	1
snake	koblo	kobo	kiongo	khôbôlô	1	1	2	1
soul	alima	alima	alima	alma	1	1	1	1
spit	kupi	kupi	kupi	tabôkha	1	1	1	2
star	xtlela	teela	têtêmbu	nteela	1	1	2	1
stick (of wood)	po	upa	po	opa	1	1	1	1
stone	budu	ubudu	buru	budu	1	1	1	1
stone of a fruit	klôsô	kôôsu	kôôthu	khôôsu	1	1	1	1
straight	dlêtu	dêêtu	deeto	dêêtu	1	1	1	1
sun	solo	isolu~usolu	tholo	sol	1	1	1	1
tail	labu	urabu	inkila	labu	1	1	2	1
tall	lôngô	lôngô	lôngô	lôngô	1	1	1	1
there	ala	la	nha	ala	1	1	1	1
they	inen	ine	ene	êney	1	1	1	1
thick (fat)	gôdô	gôdô	netaru	gôdôdô	1	1	2	1
thin (slim)	mlagu	mangu	mangu	mungulu	1	1	1	1
thou (sg.)	bô	txi	bô	bo	1	2	1	1
three	tlêxi	têêxi	têêsi	têêxi	1	1	1	1
to ask	punta	punta	puta	pongota	1	1	1	1
to bark	ladla	kupa	budja	ladala	1	2	3	1
to bite	môdê	môdê	môrê	munda	1	1	1	2
to burn	kema	kema	kema	khama	1	1	1	1
to bury	ntela	kava	tia	ntela	1	2	1	1
to come	bi	vika	bi	bi	1	2	1	1
to count	konta	konta	konta	khonta	1	1	1	1
to cultivate	ximya	ximya	ximya	ximya	1	1	1	1
to cut	kota	kota	kota	kuuta	1	1	1	1
to die	môlê	mwê	muê	môlê	1	1	1	1
to dig	kobo	kava	vina	khobo	1	2	3	2
to drink	bêbê	bêbê	bêbê	bêbê	1	1	1	1
to eat	kume	kume	mme	kumu	1	1	1	1
to fall (as in drop)	kyê	kyê	kiê	kêê	1	1	1	1
to fear	mendu	mendu	mendu	mendu	1	1	1	1
to fight	luta	luta	vugu	luta	1	1	2	1
to float	boya	boya	boya	boya	1	1	1	1
to fly	vwa	vwa	pupuka	vôa	1	1	2	1
to follow, to accompany	lêlê	lêlê	lêlê	lêlê	1	1	1	1

to give	da	da	da	da	1	1	1	1
to give birth	pali	pwe	pari	pali	1	2	1	1
to hear	têndê	têndê	êndê	têndê	1	1	1	1
to hunt (transitive)	kasa	kasa	katha	khasa	1	1	1	1
to kill	mata	mata	mata	mata	1	1	1	1
to know (a fact)	sêbê	sêbê	eta	sabê	1	1	2	1
to know a person	kônsê	kônsê	kônthê	khônsê	1	1	1	1
to laugh	li	rin	di	li	1	1	2	1
to lift up	ligi	rêgê	rigi	ligi	1	1	1	1
to live	ta	ta	ta	ta	1	1	1	1
to play	floga	foga	foga	fuuga	1	1	1	1
to pull	saya	saa	thaa	saa	1	1	1	1
to push	pinsa	pinsa	pintha	piza	1	1	1	1
to rain	suba	usuva	thuba	xiba	1	1	1	1
to request, to ask	pidji	pidi	pindji	pindji	1	1	1	1
to rub	flega	fega	fega	feega	1	1	1	1
to run	kôlê	kwê	kôlê	khôlê	1	1	1	1
to say	fla	fa	fa	faa	1	1	1	1
to scratch (an itch)	kosa	kosa	kotha	kusa	1	1	1	1
to see	bê	vê	bê	bê	1	1	1	1
to send	manda	manda	manda	manda	1	1	1	1
to sew	klôzê	kôzê	kôdhê	khôôzê	1	1	1	1
to show	musa	musa	uthua	minsia	1	1	1	1
to sing	kanta	kanta	kata	khanta	1	1	1	1
to sit	tason	tusan	tatho	tusan	1	1	1	1
to sleep	dumini	dimi	dumi	djuni	1	1	1	1
to smell (sense odor)	sela	sera	thela	sela	1	1	1	1
to split (wood)	va	va	taya	lasa	1	1	2	3
to squeeze	peta	peta	peta	peeta	1	1	1	1
to stand up	lanta	lata	lata	laanta	1	1	1	1
to steal	futa	futa	futa	fuuta	1	1	1	1
to suck	supa	supa	thupa	supa	1	1	1	1
to swell	fuma	fuma	fuma	fuma	1	1	1	1
to swim	landa	landa	landa	da	1	1	1	1
to switch, exchange	tloka	toka	toka	tuuka	1	1	1	1
to think	pensa	pensa	pentha	pensa	1	1	1	1
to throw	zuga	zuga	dhuga	zugwa	1	1	1	1
to turn (change direction)	bila	vya	bila	vla	1	1	1	1
to vomit	lansa	lansa	ngumita	luza	1	1	2	2
to walk	nda	nda	nda	nda	1	1	1	1
to want	mêsê	mêsê	mêthê	ngongo	1	1	1	2
to wash	laba	lava	laba	laba	1	1	1	1
to wet	monha	mwa	minha	mya	1	1	1	1
to whistle	supya	supya	thupya	supya	1	1	1	1
tongue	lungwa	lunge	lunga	lunge	1	1	1	1
tooth	dêntxi	idintxi	dêtxi	dêntxi	1	1	1	1
tortoise	tatalugwa	tetuga	têtêuga	tôtxiga	1	1	1	1

tree	po	upa	po	ôpa	1	1	1	1
two	dôsu	dôsu	dôthu	dôsu	1	1	1	1
underneath	basu	ubasu	kosi	basu	1	1	2	1
urine	zawa	miza	dhawa	minza	1	2	1	2
warm (as in weather)	kêntxi	kyentxi	kêtxi	kentxi	1	1	1	1
water	awa	awa	awa	awa	1	1	1	1
we	non	non	no	non	1	1	1	1
what	kêkwa	kwa	kwai	kêkhwa	1	1	1	1
when (temporal)	ola	ora	ola	ola	1	1	1	1
white	blanku	baanku	baanku	baanku	1	1	1	1
who	kêngê	ningê	ngêi	kêngê	1	2	2	1
wind (as in breeze)	ventu	uventu	ventu	oventu	1	1	1	1
wing	aza	aza	adha	aza	1	1	1	1
woman	mwala	mye	mengai	myele	1	2	3	2
wound	flida	fiida	fiira	filida	1	1	1	1
year	anu	anu	anu	anu	1	1	1	1
you (pl.)	nansê	owo	êthê	naminsêdji	1	2	3	1

ANEXO III - Lista de léxico funcional: Origem do vocabulário

Legenda
1=Português
2=Africano
3=Outro/desconhecido

Léxico Funcional							
Português		Santome		Lung'Ie		Angolar	Fa d'Ambô
Artigos							
um/uma	1	ũa	1	ũa	1	ũa	1 wan
os	2	inen	2	ine	2	ene	2 nan
as	2	inen	2	ina	2	ane	2 na
Preposições							
no/nos/na/nas	1	ni	1	na	1	ni	1 nan
por	1	plo	1	pô	1	ra	1 pol
até	1	antê	1	ten	1	atê	1 ata
com	1	ku	1	ku	1	ku	1 ku
de	1	di	1	di	1	di	1 dji
desde	1	jina	1	dina	1	zina	1 desde
para	1	pla	1	pa	1	pa	1 pa
para*	1	da	1	da	1	ra	1 da
sem	1	sê	1	xi	1	thê	1 sin
Locuções Prepositivas							
em cima	1	liba	1	ixima	1	riba	1 liba
ao lado de	1	ladu	1	zuntu	2	kota	3 lomadu
atrás de	1	tlaxi	1	taxi	2	mema	1 taaxi
debaixo de	1	basu	1	ubasu	2	kôsi	1 basu
dentro	1	glentu	1	udêntu	1	lêtu	1 dantu
meio	1	ômê	1	metedi	2	katxi	1 metadji
frente	1	wê	1	uwê	1	wê	1 deenxi
Pronomes pessoais							
1ªsg	3	n	3	n	3	n	3 am
2ªsg	1	bô	1	txi	1	bô	1 bo
3ªsg	1	ê	1	ê	1	ê	1 ê
impessoal (it/on)	2	a	2	a	2	a	2 a
1ªpl	1	non	1	non	1	no	1 non
2ªpl	2	inansê	2	owo	2	êthê	2 namsêji
3ªpl	2	inen	2	inen	2	ane	2 êneyyn
Oblíquos-átonos							
me	1	mu	1	mi	1	m	1 mun
te	2	bô	2	txi	2	bô	2 bo
Oblíquos-tônicos							
mim/comigo	3	ami	3	ami	3	m	3 (ku) amun
ti/contigo	2	bô	2	atxi	2	bô	2 txi
Reflexivo							
de mim	1	mu	1	(di) mi	1	m	1 (dôgê)-

							mun
'dele'	1	dê	1	(di) êli	1	rê	1 dêl
Possessivos							
meu/minha/meus/minhas	1	mu	1	me	1	(ri) m	1 (yi)mu
teu/tua/teus/tuas	2	bô	1	tê	2	(ri) ô	1 tua
seu/sua	1	dê	1	sê	1	(ri) rê	1 dêl
seus/suas	2	dinen	2	ine	1	(ri) rê	2 dinen
nosso/nossos	1	non	1	non	1	(ri) no	1 non
nossa/nossas	1	non	1	non	1	(ri) no	1 nosa
vosso/vossos	2	dinansê	2	owo	1	(ri)thê	3 vostulu
vossa/vossas	2	dinansê	2	owo	1	(ri)thê	3 vostala
dele/dela	1	dê	1	sun	2	(ri) ene	1 dêl
deles/delas	2	dinen	2	ine	2	(ri)ene	2 dêneyn
Demonstrativos							
este/estes/esta/estas	1	se	1	sê	1	the	1 ise
esse/esses/essa/essas	1	ise	1	isê	1	isi	1 isala
aquele/aqueles/aquela/aquelas	1	isala	1	ixila	1	isi-dha	1 isala
isto/isso/aquilo	1	isaki	1	kwisê	1	i-dhe	1 isai
Interrogativos							
que	1	kê	1	ki	1	kê	1 ke
quem	1	kêngê	1	ningê	1	ngêi	1 kêngê
qual	1	kali	1	kali	1	kutxi	1 kê-ixi
quanto	1	kantu	1	kantu	1	kantu	1 kuantu
o que	1	kwa	1	kwa	1	kwa	1 kêkhua
como	1	kuma	1	mo	1	mma	1 ama
onde	1	andji	1	kumi	1	andji	1 ana
por quê	3	punda ka manda	3	pidi kwê manda	3	kwa ma...ra	1 pokê
quando	1	kê dja	1	kê dya	3	dia kutxi	1 kê gola
numerais- cardinais							
um	1	ua	1	ua	1	ua	1 uniun
dois	1	dôsu	1	dôsu	1	dôthu	1 dôsu
três	1	tlêxi	1	têêxi	1	têêsi	1 têêxi
quatro	1	kwatlu	1	kwatu	2	kwana	1 khatulu
cinco	1	xinku	1	xinku	2	tanu	1 xinku
seis	1	sêji	1	sêy	2	thamanô	1 sêix
sete	1	sete	1	setxi	2	thambari	1 sete
oito	1	wôtô	1	wêtu	2	nake	1 ôtu
nove	1	novi	1	nove	2	uvwa	1 novi
dez	1	dexi	1	dexi	2	kwin	1 deix
Ordinais							
primeiro	1	plimê	1	pimyô	1	pumbê	1 plimêlu
Conjunções/Conectores							
e	1	ku	1	ku	1	ki	1 ku
nem	1	nê	1	ni	1	nê	1 ni
mas	1	maji	1	maji	1	mazi	1 manji
ou	1	ô	1	ô	1	o	1 ô

que	1	ku	1	ki	1	ki	1	ku
porque	3	punda	1	pôkê	3	punda	1	pokê
se	1	xi	1	xi	1	si	1	si
quando	1	ola	1	ora	1	ola	1	ola
quanto	1	kantu	1	kantu	1	kantu	1	enkwanto
então	3	so	3	so	3	thô	3	ta
depois	3	zo	3	zo	3	wo	1	despos
como	1	mo	1	mo	1	ma	3	xima-fan
do que	1	dôkê	1	dêkê	1	dôkê	1	kê
do que*	1	pasa	1	pasa	1	patha	1	pasaa
por causa de	1	plôvya	1	pôvya	3	punda	1	dajuantu
para	1	pa	1	pa	1	pa	1	pa
o que	1	kwa	1	kwa	1	kwa	1	ixi
Complementadores								
que (relativo)	1	ku	1	ki	3	ki	3	ku
que (epistémico)	1	kuma	3	ya	3	ya	3	fa
para (diretivo)	1	pa	1	pa	1	pa	1	pa
se	1	xi	1	xi	1	si	1	xi
TMA								
	3	xka	1	sa	1	tha	3	ska
	3	ka	3	ka	3	ka	3	kha
	1	tava	1	tava	1	ta	1	tan
	3	ga	3	ga	3	ga	3	ga
Partículas								
	3	an	3	a	3	an	3	a
	2	ê	2	ê	2	ê	2	ê
	2	ô	2	ô	2	ô	2	ô
Quantificadores								
todo(s) / tudo	1	tudu	1	tudu	1	turu	1	tudu
mais	1	maxi	1	maxi	1	masi	1	maixi
muito	1	montxi	1	montxi	1	montxi	1	monxi
mais do que	1	pasa	1	pasa	1	patha	1	pasaa
cada	1	kada	1	kêdê	1	kada	1	khada
um pouco	1	mina	1	minu	1	nna	1	mina
uma quantidade de	1	minda	1	minda	3	txo	1	miyida
nenhum	1	nêũa	1	nêũa	1	nêũa	3	zuguan
muito	3	yô	1	mutu	3	ndjô	1	muntu
Negação								
1º elemento da negação	1	na	1	na	1	na	1	na
2º elemento da negação	2	fa	2	fa	2	wa	2	f'
2º elemento da negação	2	fô	2	f'ô	2	fô	2	f(a)
Focalizadores								
é que	3	so	1	êli	3	thô	3	se
PM - marcas de apresentação								
eis	1	ya	1	ya	1	ya	1	yaj
Serialização verbal								
ir	1	ba	1	we	1	ba	1	bay
vir	1	bi	1	vika	1	bi	1	bi

deitar fora	1	buta	1	buta	1	dhuga ta	1	ta
dar	1	da	1	da	1	da	1	da
rodear	1	loja	1	rodya	1	taka	1	seega
pôr	1	pê	1	pwê	1	pê	1	pê
Passar	1	pasa	1	pasa	1	patha	1	pasa
sair de/vir de	1	fô	1	fo	1	fô	1	fo
tirar/tomar	1	toma	1	tama	1	tua	1	ma
virar	1	bila	1	vya	1	bila	1	vla
acompanhar/seguir	2	lêlê	2	lêlê	2	lêlê	2	lêlê

ANEXO IV - Lista do léxico funcional: Classificação de cognatos

Léxico Funcional							
Português		Santome		Lung'Ie		Angolar	Fa d'Ambô
Artigos							
um/uma	1	ũa	1	ũa	1	ũa	1 wan
os	1	inen	1	ine	1	ene	1 nan
as	1	inen	1	ina	1	ane	1 na
Preposições							
no/nos/na/nas	1	ni	1	na	1	ni	1 nan
por	1	plo	1	pô	2	ra	1 pol
até	1	antê	1	ten	1	atê	1 ata
com	1	ku	1	ku	1	ku	1 ku
de	1	di	1	di	1	di	1 dji
desde	1	jina	1	dina	1	zina	1 desde
para	1	pla	1	pa	1	pa	1 pa
para*	1	da	1	da	1	ra	1 da
sem	1	sê	1	xi	1	thê	1 sin
Locuções prepositivas							
em cima	1	liba	2	ixima	1	riba	1 liba
ao lado de	1	ladu	2	zuntu	3	kota	4 lomadu
atrás de	1	tlaxi	1	taxi	2	mema	1 taaxi
debaixo de	1	basu	1	ubasu	2	kôsi	1 basu
dentro	1	glentu	1	udêntu	1	lêtu	1 dantu
meio	1	ômê	2	metedi	3	katxi	2 metadji
frente	1	wê	1	uwê	1	wê	2 deenxi
Pronomes pessoais							
1 ^a sg	1	n	1	n	1	n	1 am
2 ^a sg	1	bô	2	txi	1	bô	1 bo
3 ^a sg	1	ê	1	ê	1	ê	1 ê
impessoal (it/on)	1	a	1	a	1	a	1 a
1 ^a pl	1	non	1	non	1	no	1 non
2 ^a pl	1	inansê	2	owo	3	êthê	1 namsêji
3 ^a pl	1	inen	1	inen	1	ane	1 êneyen
Oblíquos-átonos							
me	1	mu	1	mi	1	m	1 mun
te	1	bô	2	txi	1	bô	1 bo
Oblíquos-tônicos							
mim/comigo	1	ami	1	ami	1	m	1 (ku) amun

ti/contigo	1	bô	2	atxi	1	bô	2	txi
Reflexivo								
de mim	1	mu	1	(di) mi	1	m	1	(dôgê)-mun
'dele'	1	dê	1	(di) êli	1	rê	1	dêl
Possessivos								
meu/minha/meus/minhas	1	mu	1	me	1	(ri) m	1	(yi)mu
teu/tua/teus/tuas	1	bô	2	tê	1	(ri) ô	2	tua
seu/sua	1	dê	2	sê	1	(ri) rê	1	dêl
seus/suas	1	dinen	2	ine	3	(ri) rê	1	dinen
nosso/nossos	1	non	1	non	1	(ri) no	1	non
nossa/nossas	1	non	1	non	1	(ri) no	1	nosa
vosso/vossos	1	dinansê	2	owo	3	(ri)thê	4	vostulu
vossa/vossas	1	dinansê	2	owo	3	(ri)thê	4	vostala
dele/dela	1	dê	2	sun	3	(ri) ene	1	dêl
deles/delas	1	dinen	2	ine	2	(ri)ene	1	dêneyn
Demonstrativos								
este/estes/esta/estas	1	se	1	sê	1	the	1	ise
esse/esses/essa/essas	1	ise	1	isê	1	isi	2	isala
aquele/aqueles/aquela/aquelas	1	isala	1	ixila	1	isi-dha	1	isala
isto/isso/aquilo	1	isaki	2	kwisê	3	i-dhe	1	isai
Interrogativos								
que	1	kê	1	ki	1	kê	1	ke
quem	1	kêngê	2	ningê	2	ngêi	1	kêngê
qual	1	kali	1	kali	2	kutxi	3	kê-ixi
quanto	1	kantu	1	kantu	1	kantu	1	khuantu
o que	1	kwa	1	kwa	1	kwa	2	kêkhua
como	1	kuma	2	mo	1	mma	3	ama
onde	1	andji	2	kumi	1	andji	3	ana
por quê	1	punda ka manda	1	pidi kwê manda	2	kwa ma...ra	3	pokê
quando	1	kê dja	1	kê dya	2	dia kutxi	3	kê gola
numerais- cardinais								
um	1	ua	1	ua	1	ua	1	uniun
dois	1	dôsu	1	dôsu	1	dôthu	1	dôsu
três	1	tlêxi	1	têêxi	1	têêsi	1	têêxi
quatro	1	kwatlu	1	kwatu	1	kwana	1	khatulu
cinco	1	xinku	1	xinku	2	tanu	1	xinku
seis	1	sêji	1	sêy	2	thamanô	1	sêix
sete	1	sete	1	setxi	2	thambari	1	sete
oito	1	wôtô	1	wêtu	2	nake	1	ôtu
nove	1	novi	1	nove	2	uvwa	1	novi
dez	1	dexi	1	dexi	2	kwini	1	deix

Ordinais								
primeiro	1	plimê	1	pimyô	1	pumbê	1	plimêlu
Conjunções/Conectores								
e	1	ku	1	ku	1	ki	1	ku
nem	1	nê	1	ni	1	nê	1	ni
mas	1	maji	1	maji	1	mazi	1	manji
ou	1	ô	1	ô	1	o	1	ô
que	1	ku	1	ki	1	ki	1	ku
porque	1	punda	2	pôkê	1	punda	2	pokê
se	1	xi	1	xi	1	si	1	si
quando	1	ola	1	ora	1	ola	1	ola
quanto	1	kantu	1	kantu	1	kantu	2	enkwanto
então	1	so	1	so	1	thô	2	ta
depois	1	zo	1	zo	1	wo	2	despos
como	1	mo	1	mo	1	ma	2	xima-fan
do que	1	dôkê	1	dêkê	1	dôkê	1	kê
do que*	1	pasa	1	pasa	1	patha	1	pasaa
por causa de	1	plôvya	1	pôvya	2	punda	3	dajuantu
para	1	pa	1	pa	1	pa	1	pa
o que	1	kwa	1	kwa	1	kwa	2	ixi
Complementadores								
que (relativo)	1	ku	1	ki	1	ki	1	ku
que (epistémico)	1	kuma	2	ya	2	ya	3	fa
para (diretivo)	1	pa	1	pa	1	pa	1	pa
se	1	xi	1	xi	1	si	1	xi
TMA								
	1	xka	2	sa	2	tha	1	ska
	1	ka	1	ka	1	ka	1	kha
	1	tava	1	tava	1	ta	1	tan
	1	ga	1	ga	1	ga	1	ga
Partículas finais								
	1	an	1	a	1	an	1	a
	1	ê	1	ê	1	ê	1	ê
	1	ô	1	ô	1	ô	1	ô
Quantificadores								
todo(s) / tudo	1	tudu	1	tudu	1	turu	1	tudu
mais	1	maxi	1	maxi	1	masi	1	maixi
muito	1	montxi	1	montxi	1	montxi	1	monxi
mais do que	1	pasa	1	pasa	1	patha	1	pasaa
cada	1	kada	1	kêdê	1	kada	1	khada
um pouco	1	mina	1	minu	1	nna	1	mina

uma quantidade de	1	minda	1	minda	2	txo	1	miyida
nenhum	1	nêuan	1	nêuan	1	nêuan	2	zuguan
muito	1	yô	2	mutu	1	ndjô	2	muntu
Negação								
1º elemento da negação	1	na	1	na	1	na	1	na
2º elemento da negação	1	fa	1	fa	1	wa	1	f'
2º elemento da negação	1	fô	1	fô	1	fô	1	f(a)
Focalizadores								
é que	1	so	2	êli	1	thô	1	se
PM - marcas de apresentação								
eis	1	ya	1	ya	1	ya	1	yaj
Serialização verbal								
ir	1	ba	2	we	1	ba	1	bay
vir	1	bi	2	vika	1	bi	1	bi
deitar fora	1	buta	1	buta	2	dhuga ta	3	ta
dar	1	da	1	da	1	da	1	da
rodear	1	loja	1	rodya	2	taka	3	seega
pôr	1	pê	1	pwê	1	pê	1	pê
passar	1	pasa	1	pasa	1	patha	1	pasa
sair de/vir de	1	fô	1	fo	1	fô	1	fo
tirar/tomar	1	toma	1	tama	1	tua	2	ma
virar	1	bila	1	vya	1	bila	1	vla
acompanhar/seguir	1	lêlê	1	lêlê	1	lêlê	1	lêlê

ANEXO V – Lista de traços sintáticos: Classificados

	Traços	ST	LI	ANG	FA
1	qual a ordem básica dos constituintes [1]				
	S-V-O	+	+	+	+
	V-O-S				
	V-S-O				
	O-V-S				
	O-S-V				
2	qual a ordem entre o SN sujeito possessivo(possuidor) e o SN argumento possessivo (possuído) [2]				
	SN sujeito possessivo(possuidor) - SN argumento possessivo (possuído)				
	SN argumento possessivo (possuído)-SN sujeito possessivo(possuidor)	+	+	+	+
3	qual a posição relativa entre pronomes possessivos e nomes:				
	pronome possessivo-nome				
	nome-pronome possessivo	+	+	+	+
4	qual a ordem canónica entre os adjetivos e o nome [3]				
	nome-adjectivo	+	+	+	+
	adjectivo-nome				
5	os adjetivos podem ocorrer em posição:				
	pré-nominal	+	+	+	+
	pós-nominal	+	+	+	+
6	qual a ordem entre a adposição e o SN [4]				
	adposição-nome	+	+	+	+
	nome-adposição				
7	qual a ordem entre o demonstrativo e o nome [5]				
	demonstrativo-nome				
	nome-demonstrativo	+	+	+	+
8	qual a posição relativa entre demonstrativos e possessivos:				
	demonstrativo-possessivo	+	+	+	
	possessivo-demonstrativo		+		+
9	qual a posição canónica dos pronomes demonstrativos em relação às orações relativas				
	NP- DEM- OR	+	+		+
	NP- OR- DEM			+	
10	qual a ordem canónica entre os numerais cardinais e o nome [6]				
	numeral-nome	+	+	+	
	nome-numeral				+
11	qual a ordem entre a oração relativa e o nome [7]				
	oração relativa-nome				
	nome-oração relativa	+	+	+	+
12	qual a ordem canónica entre os advérbios de intensidade e os adjetivos [8]				

	advérbios de intensidade-adjectivos		+	+	
	adjectivos-advérbios de intensidade	+		+	+
13	qual a posição do artigo definido no sintagma nominal [9]				
	artigo definido-nome	+			+
	nome-artigo definido				
	a língua não possui artigos definidos		+	+	
14	qual a posição do artigo indefinido no sintagma nominal [10]				
	artigo indefinido-nome	+		+	+
	nome-artigo indefinido		+		
	a língua não possui artigos indefinidos				
15	qual a ordem canónica entre os advérbios de frequência, o verbo e o objeto [11]				
	advérbio de frequência-verbo-objecto	+	+	+	+
	verbo-objecto-advérbio de frequência	+	+	+	+
	verbo-advérbio de frequência-objecto		+	+	
16	qual a posição canónica do pronome interrogativo? [12]				
	posição inicial	+	+	+	+
	posição não inicial				
17	os pronomes interrogativos -wh são constituídos por: [19]				
	palavras monomorfémicas				
	uma expressão composta		+	+	
	duas expressões compostas	+			+
18	a língua tem um elemento interrogativo final que interroga DP's:				
	sim	+	+	+	+
	não				
19	qual a posição que ocupa o marcador de plural em relação ao nome [23]				
	pré- nominal*	+	+	+	+
	pós-nominal				
	a língua não possui um marcador nominal de plural**				
20	em que posição ocorrem os marcadores de plural dentro do sintagma nominal:				
	pré-nominal	+		+	
	pós-nominal		+		+
21	os demonstrativos podem ter a mesma função que os artigos definidos [28]				
	artigos definido distintos dos demonstrativos	+			+
	a língua não possui artigos definidos		+	+	
22	o artigo indefinido deriva do numeral 'um' [29]				
	sim	+	+	+	
	não				+
23	há coocorrência entre o demonstrativo e o artigo definido? [31]				
	sim	+			+
	não				
	a língua não possui artigos definidos		+	+	

24	os demonstrativos pronominais e adnominais apresentam a mesma forma? [32]				
	sim				
	não	+	+	+	+
25	a língua possui um determinante demonstrativo que é, por norma, seguido de um segundo determinante demonstrativo				
	sim			+	
	não	+	+		+
26	a língua pluraliza o demonstrativo				
	sim				+
	não	+	+	+	
27	todos os ordinais derivam diretamente dos cardinais? [35]				
	os ordinais 'primeiro', 'segundo' e alguns mais são supletivos, os restantes derivam dos cardinais.	+	+		
	ordinais e cardinais são idênticos, exceto no primeiro/um			+	+
28	de que forma é feita a marcação de possuidores pronominais? [37]				
	preposição-pronome possessivo	+			
	pronome possessivo-preposição		+	+	+
29	numa construção possessiva adnominal de que forma é feita a marcação de caso possessivo? [38]				
	através de um marcador adposicional ou de caso possessivo	+			
	não há marcação	+	+	+	+
30	de que forma é expresso o possuidor nominal quando ocorre de forma independente? [39]				
	preposição 'especial' + possuidor	+		+	+
	palavra 'especial' + possuidor		+		+
31	há concordância em género entre o adjetivo e o nome? [40]				
	há concordância em alguns adjetivos		+		
	não há concordância	+		+	+
32	nas construções comparativas o adjetivo é marcado? [41]				
	sim	+	+	+	+
	não				+
33	nas construções comparativas o padrão é marcado? [42]				
	sim	+	+		+
	não			+	
34	qual a posição de TMA em relação ao verbo [43]				
	imediatamente antes do verbo	+		+	+
	à esquerda do verbo	+	+	+	
35	qual a ordem interna entre os marcadores de tempo, modo e aspeto [44]				
	tempo-modo-aspecto				
	tempo-aspecto-modo				

	modo-tempo-aspecto	+	+		
	não se aplica			+	+
36	que elementos podem ocorrer entre o marcador de passado e o verbo? [45]				
	determinados itens lexicais	+	+		
	apenas marcadores gramaticais			+	+
37	que elementos podem ocorrer entre o marcador progressivo e o verbo? [46]				
	determinados itens lexicais		+	+	
	nenhum elemento pode ocorrer	+			+
38	que outras funções tem o marcador progressivo? [47]				
	função de marcador progressivo e habitual				+
	apenas função de marcador progressivo	+	+	+	
39	que outras funções tem o marcador habitual? [48]				
	função de marcador habitual, presente e futuro	+	+	+	
	apenas função de marcador habitual				+
40	qual o tipo de sistema temporal e aspetual [49]				
	sistema puramente aspetual				
	sistema puramente temporal				
	sistema temporal-aspetual misto	+	+	+	+
41	em orações negativas, o sistema de TMA ocorre em forma: [50]				
	reduzida		+		
	padrão	+		+	+
	diferente				
42	de que forma é assinalada a ditransitividade com as construções 'give': [60]				
	construções de duplo objeto	+	+	+	
	construções de objeto indireto				+
43	em construções ditransitivas, qual a ordem entre o recipiente e o tema objeto: [61]				
	sujeito-verbo-recipientes-tema	+	+	+	
	sujeito-verbo-tema-recipientes				+
44	de que forma o sujeito é expresso quando se refere ao locutor, interlocutor ou a outro referente [62]				
	uso de pronome obrigatório	+	+	+	+
	uso de pronome opcional				

45	nas construções do tipo 'parecer' é usado um sujeito expetivo [63]				
	sim		+		
	não		+		
	não há construções do tipo 'parecer'	+		+	+
46	nas construções existenciais é usado um sujeito expetivo [64]				
	sim	+	+		+
	não		+	+	
47	de que forma são expressos os verbos meteorológicos [65]				
	ele chove'	+			
	a chuva chove'	+	+	+	
	a chuva cai'		+	+	+
48	com' comitativo e instrumental são expressos da mesma forma? [70]				
	sim	+	+	+	+
	não				
49	a língua usa o mesmo coordenador para sintagmas nominais e comitativos [71]				
	sim	+	+	+	+
	não				
50	conjunção nominal e conjunção verbal são expressos da mesma forma [72]				
	sim				
	não	+	+	+	+
51	o verbo copulativo ocorre em orações com sintagmas nominais predicativos, com função de inclusão de classe [73]				
	sim, com cópula variável		+	+	+
	sim, com cópula invariável	+			
	não				
52	o verbo copulativo ocorre em orações com adjetivos predicativos [74]				
	sim, com cópula variável	+	+	+	+
	sim, com cópula invariável				
	não				
53	o verbo copulativo ocorre em orações com sintagmas predicativos locativos? [75]				
	sim, com cópula variável		+		
	sim, com cópula invariável	+		+	+
	não				

54	de que forma os sintagmas nominais predicativos e os sintagmas predicativos locativos são interpretados [76]				
	identidade	+			+
	sobreposição			+	
	diferenciação		+		
55	de que forma os predicativos possessivos são expressos: [77]				
	transitivo	+	+		+
	comitativo	+	+	+	+
	tópico				+
56	o verbo existencial é semelhante ao verbo transitivo de posse 'ter' [78]				
	identidade		+		
	sobreposição			+	+
	diferenciação	+			
57	de que forma é expressa a transitividade- 'ir para': [79]				
	preposição				
	não apresenta marcador de caso (para além do verbo)	+	+	+	+
58	de que forma é expressa a transitividade- 'vir de': [80]				
	preposição				
	verbos seriais	+	+	+	+
59	qual a estratégia usada para expressar as orientações 'ir para' e 'vir de' [81]				
	identidade				
	sobreposição				
	diferenciação	+	+	+	+
60	de que forma é expressa a transitividade em verbos do tipo 'empurrar' [82]				
	construções de verbos seriais	+	+	+	+
	construções de verbos seriais seguido de preposição				
61	de que forma é expressa a transitividade em verbos do tipo 'puxar' [83]				
	at-rest' é usado para expressar direcionalidade 'vir de'	+			
	construções de verbos seriais	+	+	+	+
62	segundo elemento da construção serial especifica a direção da ação a que se refere o primeiro verbo [84]				
	existência dos direcionais 'ir' e 'vir'	+	+	+	+
	não existência dos direcionais 'ir' e 'vir'				

63	take serials: qual o papel do segundo elemento da construção serial [85]				
	instrumento do segundo verbo, com pronome resumptivo	+	+	+	
	instrumento do segundo verbo, sem pronome resumptivo	+	+	+	
	tema do segundo verbo, possível interpretação literal	+	+	+	+
	tema do segundo verbo, impossível interpretação literal	+	+	+	+
64	give serials: qual o papel do segundo elemento da construção serial [86]				
	give' na segunda posição, recipiente ou beneficiário	+	+	+	
	give' na segunda posição, apenas recipiente				+
65	as construções reflexivas são expressas através de: [87]				
	expressões implícitas	+			+
	pronomes reflexivos com corpo ou parte do corpo	+	+	+	+
	pronome reflexivos compostos com intensificador		+		
66	qual a relação entre os intensificadores e os reflexivos [88]				
	intensificadores e reflexivos são diferenciados	+		+	+
	intensificadores e reflexivos sobrepõem-se		+		
67	qual a relação entre os marcadores de reciprocidade e os reflexivos [89]				
	construções de reciprocidade especiais com base na expressão 'outro'	+	+	+	
	outro tipo de construção recíproca				+
68	de que forma são expressas as construções passivas [90]				
	passivas sem expressão verbal		+	+	
	outras construções passivas atípicas				+
	ausência de construções passivas	+			
69	de que forma é introduzida a oração relativa de sujeito [92]				
	partículas relativas e lacuna (gap)	+	+		
	pronome relativo			+	
	zero e pronome resumptivo				+
70	de que forma é introduzida a oração relativa de objeto [93]				
	partícula relativa e lacuna (gap)	+	+		
	pronome relativo			+	+
	zero e lacuna (gap)				+
71	de que forma é introduzida a oração relativa instrumento [94]				
	partícula relativa e pronome resumptivo	+	+		
	pronome relativo com pronome resumptivo			+	+

	significado implícito		+		
72	em construções de discurso direto, verbos do tipo 'dizer' [95]				
	complementador não relacionado com verbo sincronicamente	+	+	+	
	complementador idêntico ao verbo 'dizer'			+	+
	sem complementador	+	+	+	
	complemento consiste em verbo 'dizer' + outro marcador			+	
73	qual o complementado usado com verbos do tipo 'saber' [96]				
	complementador não relacionado com verbo sincronicamente	+	+	+	
	complementador idêntico ao verbo 'saber'				+
	sem complementador				
74	de que forma é expresso o sujeito da oração complemento do verbo 'querer' [97]				
	o sujeito do complemento está implícito à esquerda	+	+	+	+
	o sujeito do complemento é expresso abertamente	+		+	
75	para os verbos do tipo 'saber' e 'querer' o complementador é: [98]				
	diferente	+	+	+	+
	semelhante				
76	construções de duplicação verbal são possíveis em orações temporais [99]				
	sim				
	não	+	+	+	+
77	qual o tipo de morfemas negativos [100]				
	marcador negativo bipartido	+		+	+
	partícula negativa		+		
78	qual a posição do marcador de negação em relação ao verbo [101]				
	bipartido, antes do verbo e depois do objeto	+		+	+
	depois do verbo e do seu objeto		+		
79	o pronome indefinido coocorre com o predador de negação [102]				
	sim	+	+	+	+
	não				
80	nas orações finais, o elemento de negação final :				
	ocorre				
	não ocorre	+	+	+	+
81	na negação de constituintes, ocorre apenas o elemento de negação final				
	sim	+	+	+	+

	não				
82	de que forma são expressas as questões polares [103]				
	através da entoação interrogativa	+	+	+	+
	através de marcadores interrogativos finais				
83	de que forma é focalizado o sintagma nominal [104]				
	ocorrência de partícula após o foco	+		+	+
	clivagem com partícula de focalização após o foco		+	+	
	clivagem 'bare'		+		
84	há duplicação verbal de forma a focalizar a situação referida pelo verbo [105]				
	verbo ocorre com cópia na segunda oração	+	+	+	+
	não há duplicação				
85	a partícula de focalização 'também' ocorre [106]				
	antes do elemento focalizado				
	depois do elemento focalizado	+	+	+	+

ANEXO VI - Lista de traços sintáticos – Versão para *SplitsTree4*

	ST	LI	ANG	FA
1 qual a ordem básica dos constituintes [1]	1	1	1	1
2 qual a ordem entre o SN sujeito possessivo(possuidor) e o SN argumento possessivo (possuído) [2]	1	1	1	1
3 qual a posição relativa entre pronomes possessivos e nomes:	1	1	1	1
4 qual a ordem canónica entre os adjetivos e o nome [3]	1	1	1	1
5 os adjetivos podem ocorrer em posição:	1	1	1	1
6 qual a ordem entre a adposição e o SN [4]	1	1	1	1
7 qual a ordem entre o demonstrativo e o nome [5]	1	1	1	1
8 qual a posição relativa entre demonstrativos e possessivos:	1	2	1	3
9 qual a posição canónica dos pronomes demonstrativos em relação às orações relativas	1	1	2	1
10 qual a ordem canónica entre os numerais cardinais e o nome [6]	1	1	1	2
11 qual a ordem entre a oração relativa e o nome [7]	1	1	1	1
12 qual a ordem canónica entre os advérbios de intensidade e os adjetivos [8]	1	2	3	1
13 qual a posição do artigo definido no sintagma nominal [9]	1	2	2	1
14 qual a posição do artigo indefinido no sintagma nominal [10]	1	2	1	1
15 qual a ordem canónica entre os advérbios de frequência, o verbo e o objeto [11]	1	2	2	1
16 qual a posição canónica do pronome interrogativo? [12]	1	1	1	1
17 os pronomes interrogativos -wh são constituídos por: [19]	1	2	2	1
18 a língua tem um elemento interrogativo final que interroga DP's:	1	1	1	1
19 qual a posição que ocupa o marcador de plural em relação ao nome [23]	1	1	1	1
20 em que posição ocorrem os marcadores de plural dentro do sintagma nominal:	1	2	1	2
21 os demonstrativos podem ter a mesma função que os artigos definidos [28]	1	2	2	1
22 o artigo indefinido deriva do numeral 'um' [29]	1	1	1	2
23 há coocorrência entre o demonstrativo e o artigo definido? [31]	1	2	2	1
24 os demonstrativos pronominais e adnominais apresentam a mesma forma? [32]	1	1	1	1
25 a língua possui um determinante demonstrativo que é, por norma, seguido de outro determinante demonstrativo	1	1	2	1
26 a língua pluraliza o demonstrativo	1	1	1	2
27 todos os ordinais derivam diretamente dos cardinais? [35]	1	1	2	2
28 de que forma é feita a marcação de possuidores pronominais? [37]	1	2	2	2
29 numa construção possessiva adnominal de que forma é feita a marcação de caso possessivo? [38]	1	2	2	2
30 de que forma é expresso o possuidor nominal quando ocorre de forma independente? [39]	1	2	1	3
31 há concordância em género entre o adjetivo e o nome? [40]	1	2	1	1
32 nas construções comparativas o adjetivo é marcado? [41]	1	1	1	2
33 nas construções comparativas o padrão é marcado? [42]	1	1	2	1

34	qual a posição de TMA em relação ao verbo [43]	1	2	1	3
35	qual a ordem interna entre os marcadores de tempo, modo e aspeto [44]	1	1	2	2
36	que elementos podem ocorrer entre o marcador de passado e o verbo? [45]	1	1	2	2
37	que elementos podem ocorrer entre o marcador progressivo e o verbo? [46]	1	2	2	1
38	que outras funções tem o marcador progressivo? [47]	1	1	1	2
39	que outras funções tem o marcador habitual? [48]	1	1	1	2
40	qual o tipo de sistema temporal e aspetual [49]	1	1	1	1
41	em orações negativas, o sistema de TMA ocorre em forma: [50]	1	2	1	1
42	de que forma é assinalada a ditransitividade com as construções 'give': [60]	1	1	1	2
43	em construções ditransitivas, qual a ordem entre o recipiente e o tema objeto: [61]	1	1	1	2
44	de que forma o sujeito é expesso quando se refere ao locutor, interlocutor ou a outro referente [62]	1	1	1	1
45	nas construções do tipo 'parecer' é usado um sujeito expetivo [63]	1	2	1	1
46	nas construções existenciais é usado um sujeito expetivo [64]	1	2	3	1
47	de que forma são expressos os verbos meteorológicos [65]	1	2	2	3
48	com' comitativo e instrumental são expressos da mesma forma? [70]	1	1	1	1
49	a língua usa o mesmo coordenador para sintagmas nominais e comitativos [71]	1	1	1	1
50	conjunção nominal e conjunção verbal são expressos da mesma forma [72]	1	1	1	1
51	o verbo copulativo ocorre em orações com sintagmas nominais predicativos, com função de inclusão de classe [73]	1	2	2	2
52	o verbo copulativo ocorre em orações com adjetivos predicativos [74]	1	1	1	1
53	o verbo copulativo ocorre em orações com sintagmas predicativos locativos? [75]	1	2	1	1
54	de que forma os sintagmas nominais predicativos e os sintagmas predicativos locativos são interpretados [76]	1	2	3	1
55	de que forma os predicativos possessivos são expressos: [77]	1	1	2	3
56	o verbo existencial é semelhante ao verbo transitivo de posse 'ter' [78]	1	2	3	3
57	de que forma é expressa a transitividade- 'ir para': [79]	1	1	1	1
58	de que forma é expressa a transitividade- 'vir de': [80]	1	1	1	1
59	qual a estratégia usada para expressar as orientações 'ir para' e 'vir de' [81]	1	1	1	1
60	de que forma é expressa a transitividade em verbos do tipo 'empurrar' [82]	1	1	1	1
61	de que forma é expressa a transitividade em verbos do tipo 'puxar' [83]	1	2	2	2
62	segundo elemento da construção serial especifica a direção da ação a que se refere o primeiro verbo [84]	1	1	1	1

63	take serials: qual o papel do segundo elemento da construção serial [85]	1	1	1	2
64	give serials: qual o papel do segundo elemento da construção serial [86]	1	1	1	2
65	as construções reflexivas são expressas através de: [87]	1	2	3	1
66	qual a relação entre os intensificadores e os reflexivos [88]	1	2	1	1
67	qual a relação entre os marcadores de reciprocidade e os reflexivos [89]	1	1	1	2
68	de que forma são expressas as construções passivas [90]	1	2	2	3
69	de que forma é introduzida a oração relativa de sujeito [92]	1	1	2	3
70	de que forma é introduzida a oração relativa de objeto [93]	1	1	2	3
71	de que forma é introduzida a oração relativa instrumento [94]	1	2	3	3
72	em construções de discurso direto, verbos do tipo 'dizer' [95]	1	1	2	3
73	qual o complementado usado com verbos do tipo 'saber' [96]	1	1	1	2
74	de que forma é expresso o sujeito da oração complemento do verbo 'querer' [97]	1	2	1	2
75	para os verbos do tipo 'saber' e 'querer' o complementador é: [98]	1	1	1	1
76	construções de duplicação verbal são possíveis em orações temporais [99]	1	1	1	1
77	qual o tipo de morfemas negativos [100]	1	2	1	1
78	qual a posição do marcador de negação em relação ao verbo [101]	1	2	1	1
79	o pronome indefinido coocorre com o predador de negação [102]	1	1	1	1
80	nas orações finais, o elemento de negação final :	1	1	1	1
81	na negação de constituintes, ocorre apenas o elemento de negação final	1	1	1	1
82	de que forma são expressas as questões polares [103]	1	1	1	1
83	de que forma é focalizado o sintagma nominal [104]	1	2	3	1
84	há duplicação verbal de forma a focalizar a situação referida pelo verbo [105]	1	1	1	1
85	a partícula de focalização 'também' ocorre [106]	1	1	1	1